

METHODS AND COMPOSITIONS FOR DIAGNOSING AND TREATING RHEUMATOID ARTHRITIS

5

Cross-Reference to Related Applications

This application claims the benefit of U.S. Provisional Application No. 60/255,861, filed December 15, 2000, the contents of which are specifically incorporated by reference herein.

10 Background of the Invention

Inflammatory reactions are the cause of a significant number of diseases or disorders, some of which lack appropriate methods of treatment. For example, rheumatoid arthritis (R.A.) is a systematic inflammatory disease that commonly affects the joints, particularly those of the hands and feet. The onset of rheumatoid arthritis can occur slowly, ranging from a few weeks to a few months, or the condition can surface rapidly in an acute manner.

Today, over 2,500,000 individuals are diagnosed with rheumatoid arthritis in the United States alone (1% of population), with some statistics indicating from 6.5 to 8 million potentially afflicted with the disease. Women are affected 2-3 times more often than men. The disease can occur at any age and typically will increase in incidence with age.

The classic early symptoms of rheumatoid arthritis include stiffness, tenderness, fever, subcutaneous nodules, achy joints, and fatigue. The joints of the hands, feet, knees and wrists are most commonly affected, with eventual involvement of the hips, elbows and shoulders. As the joints stiffen and swell, any type of motion becomes very painful and difficult. The more severe cases of rheumatoid arthritis can lead to intense pain and eventual joint destruction. Some 300,000 bone and joint replacement surgical procedures are performed annually in an effort to alleviate the pain and mobility loss resultant from arthritis related joint destruction.

The effective treatment of rheumatoid arthritis has generally comprised a combination of medication, exercise, rest and proper joint protection therapy. The therapy for a particular patient depends on the severity of the disease and the joints that are involved. Aspirin is widely used for pain and to reduce inflammation. In addition to aspirin, non-steroidal anti-inflammatory drugs, corti-costeroids, gold salts, anti-malarials and systemic immunosuppressants are widely used in

moderate to advanced cases. The use of steroids and immunosuppressants, however, has significant risks and side effects both in terms of toxicity and vulnerability to potentially lethal conditions.

There, thus exists a need for methods of diagnosing and treating inflammatory diseases, e.g., rheumatoid arthritis, which do not entail the potentially lethal side effects associated with the treatments described above.

Summary of the Invention

In one embodiment, the invention provides diagnostic methods, composition and devices for monitoring and/or predicting the existence, development and/or evolution of R.A. in a subject. Preferred methods comprise determining levels of expression of one or more genes characteristic of R.A. in a cell and comparing these to the levels of expression of these genes in other cells.

Comparison of the expression levels can be performed visually. In a preferred embodiment, the comparison is performed by a computer. In one embodiment, expression levels of genes characteristic of R.A. in cells of subjects having R.A. are stored in a computer. The computer may optionally comprise expression levels of these genes in normal cells. The data representing expression levels of the genes in a patient being diagnosed are then entered into the computer, and compared with one or more of the expression levels stored in the computer. The computer calculates differences and presents data showing the differences in expression of the genes in the two types of cells.

Accordingly, in one embodiment, the invention provides computer-readable media comprising a plurality of digitally encoded values representing the levels of expression of a plurality of genes which are up- or down-regulated in R.A. in a cell characteristic of R.A. In one embodiment, a computer-readable medium includes values representing levels of expression of one or more genes encoding kinases, phosphatases or genes which are located on chromosome 6, region p21.3, such as those highlighted in the Tables. In another embodiment, the computer-readable medium comprises values of levels of expression of a plurality of genes selected from the group consisting of SOCS3 (CISH3); RAGE (AGER); LST-1 (LY117); serum amyloid (SAA) 1-3; HMG-1; S100 A8, A9, and A12; Secretory Leukocyte Protease Inhibitor (SLPI); glucocorticoid leucine zipper (GILZ); PTPN-18; GADD-45A and B; Legumain (PRSC1);

follistatin-like 1 (FST1); lipocalin 2 (Lcn2); glucose phosphate isomerase (GPI); Serine Protease Inhibitor (SpiL); and TSG-6. In a preferred embodiment, a computer-readable medium comprises values representing levels of expression of at least 5 of these genes. In another embodiment, a computer-readable medium comprises the levels of expression of at least 10 genes characteristic of R.A. in a cell characteristic of R.A. A computer-readable medium may also comprise values representing levels of expression of at least 50% of the genes set forth in Tables 1-5. Optionally, a computer-readable medium further comprises at least one value representing a level of expression of at least one gene characteristic of R.A. in a normal counterpart cell. The values on a computer-readable medium may represent ratios of, or differences between, a level of expression of a gene characteristic of R.A. in a cell characteristic of R.A. and a level of expression of the gene in a normal counterpart cell. In a preferred embodiment, less than about 50% of the values on the computer-readable medium represent expression levels of genes which are not characteristic of R.A.

The invention also provides computer systems, comprising a database comprising values representing expression levels of a plurality of genes which are up- or down-regulated in R.A., and including, e.g., one or more genes highlighted or marked with a star in the Tables, e.g., one or a plurality of genes selected from the group consisting of SOCS3 (CISH3); RAGE (AGER); LST-1 (LY117); SAA 1-3; HMG-1; S100 A8, A9, and A12; SLPI; GILZ; PTPN-18; GADD-45A and B; Legumain (PRSC1); FST1; Lcn2; GPI; SpiL; and TSG-6, in a cell characteristic of R.A.; and, a processor having instructions to, (i) receive at least one query value representing at least one level of expression of at least one gene represented in the database, and, (ii) compare the at least one query value and the at least one database value. The instructions to receive may include instructions to provide a user interface. The instructions may further include instructions to display at least one comparison and/or to create at least one record based on the comparison. The computer system may further including instructions to display the at least one record.

Also provided by the invention are computer programs for analyzing levels of expression of a plurality of genes characteristic of R.A. in a cell, the computer program being disposed on a computer readable medium and including instructions for causing a processor to (i) receive query values representing levels of expression of a plurality of genes characteristic of R.A. in a cell, and, (ii) compare the query values with levels of expression of the plurality of genes in a cell characteristic of R.A. The computer program may further comprise instructions to display at

least one comparison. The instructions to compare may include instructions to retrieve the at least one level expression value from a computer readable medium and/or from a database. The instructions to receive may include instructions to provide a user interface.

In another embodiment, the invention provides computer programs for analyzing an expression profile of a cell characteristic of R.A. in a subject, the computer programs being disposed on a computer readable medium and including instructions for causing a processor to (i) receive at least one query expression profiles comprising a plurality of values, each value representing a level of expression of a gene characteristic of R.A. in a cell characteristic of R.A., and, (ii) compare the at least one query expression profile and at least one reference expression profile comprising a plurality of values, each value representing a level of expression of a gene characteristic of R.A. in a particular cell.

Also within the scope of the invention are compositions, such as compositions comprising a plurality of detection agents of genes which are up- or down-regulated in R.A., e.g., one or more genes highlighted or marked with a star in the Tables, e.g., one or a plurality of genes selected from the group consisting of SOCS3 (CISH3); RAGE (AGER); LST-1 (LY117); SAA 1-3; HMG-1; S100 A8, A9, and A12; SLPI; GILZ; PTPN-18; GADD-45A and B; Legumain (PRSC1); FST1; Lcn2; GPI; SpiL; and TSG-6, which are capable of detecting the expression of the genes or the polypeptide encoded by the genes, and wherein less than about 50% of the detection agents are genes which are not characteristic of R.A. The detection agents may be isolated nucleic acids which hybridize specifically to nucleic acids corresponding to the genes. Compositions may comprise isolated nucleic acids which hybridize specifically to at least five genes selected from the group consisting of SOCS3 (CISH3); RAGE (AGER); LST-1 (LY117); SAA 1-3; HMG-1; S100 A8, A9, and A12; SLPI; GILZ; PTPN-18; GADD-45A and B; Legumain (PRSC1); FST1; Lcn2; GPI; SpiL; and TSG-6. In another embodiment, a composition may comprise isolated nucleic acids which hybridize specifically to at least 10 or 100 different genes characteristic of R.A. The detection agents may also detect the polypeptides encoded by the genes and may be, e.g., antibodies.

The invention also provides solid surfaces to which are linked a plurality of detection agents of genes which are up- or down-regulated in R.A., e.g., one or more genes highlighted or marked with a star in the Tables, e.g., one or a plurality of genes selected from the group consisting of SOCS3 (CISH3); RAGE (AGER); LST-1 (LY117); SAA 1-3; HMG-1; S100 A8,

A9, and A12; SLPI; GILZ; PTPN-18; GADD-45A and B; Legumain (PRSC1); FST1; Lcn2; GPI; SpiL; and TSG-6, which detection agents are capable of detecting the expression of the genes or the polypeptide encoded by the genes, and wherein less than about 50% of the detection agents on the solid surface are not detecting genes characteristic of R.A. The detection agents
5 may be isolated nucleic acids which hybridize specifically to the genes. The detection agents may be covalently linked to the solid surface.

Other compositions provided by the invention include compositions, such as pharmaceutical compositions comprising agonists or antagonists of a plurality of genes characteristic of R.A., such as antagonists of one or a plurality of genes selected from the group
10 consisting of SOCS3 (CISH3); RAGE (AGER); LST-1 (LY117); SAA 1-3; HMG-1; S100 A8, A9, and A12; SLPI; GILZ; PTPN-18; GADD-45A and B; Legumain (PRSC1); FST1; Lcn2; GPI; SpiL; and TSG-6. Agonists may be polypeptides encoded by the genes or functional fragments or equivalents thereof, which may be fused to a transcytosis polypeptide. Agonists may also be genes encoding the polypeptides and the nucleic acids may be in one or more expression vectors. Antagonists may be antisense nucleic acids, siRNAs, ribozymes or dominant negative mutants.

The invention provides methods for determining the difference between levels of expression of a one or a plurality of genes characteristic of R.A. in a cell and reference levels of expression of the genes, comprising (i) providing RNA from a cell; (ii) determining levels of
15 RNA of a plurality of genes which are up- or down-regulated in R.A., e.g., one or more genes highlighted or marked with a star in the Tables, e.g., one or a plurality of genes selected from the group consisting of SOCS3 (CISH3); RAGE (AGER); LST-1 (LY117); SAA 1-3; HMG-1; S100 A8, A9, and A12; SLPI; GILZ; PTPN-18; GADD-45A and B; Legumain (PRSC1); FST1; Lcn2; GPI; SpiL; and TSG-6 to obtain the levels of expression of the plurality
20 of genes in the cell; and (iii) comparing the levels of expression of the plurality of genes in the cell to a set of reference levels of expression of the genes, to thereby determine the difference between levels of expression of the plurality of genes characteristic of R.A. in the cell and reference levels of expression of the genes. The set of reference levels of expression may include the levels of expression of the genes in a subject having R.A. The set of reference levels
25 of expression may further include the levels of expression of the genes in a subject who does not have R.A. The method may comprise incubating a nucleic acid sample derived from the RNA of

the cell of the subject with nucleic acids corresponding to the genes, under conditions wherein two complementary nucleic acids hybridize to each other. The nucleic acids corresponding to the genes may be attached to a solid surface. The method may comprise entering the levels of expression of the plurality of genes into a computer which comprises a memory with values representing the set of reference levels of expression. Comparing the level may comprise providing computer instructions to perform.

The invention provides methods for determining whether a subject has or is likely to develop R.A., comprising obtaining a cell from the subject and comparing gene expression levels in the cell to those of a set of reference levels of expression, e.g., as described above, wherein similar levels of expression of the plurality of genes indicates that the subject has or is likely to develop R.A. In a preferred embodiment, the cell is a peripheral blood mononuclear cell (PBMC) and the set of reference levels of expression includes the levels of expression of the genes in a PBMC of a subject having R.A. The cell may be a PBMC and the set of reference levels of expression includes the average of levels of expression of the genes in a PBMC of a plurality of subjects having R.A. The method may further comprising iteratively providing RNA from the subject and determining the level of RNA, such as to determine an evolution of the levels of expression of the genes in the subject.

Also within the scope of the invention are methods for determining whether a therapy for R.A. is effective in a subject having R.A. who is receiving the therapy. In a exemplary embodiment, the method comprises obtaining a cell from the subject and comparing levels of expression in the cell of the subject to those in subjects having R.A. and in subjects who do not have R.A., e.g., as described above, wherein levels of expression in the cell of the subject that are more similar to those of the subject having R.A. than the subject who does not have R.A. indicates that the therapy is not effective, whereas levels of expression in the cell of the subject that are more similar to those of the subject not having R.A. than the subject having R.A. indicates that the therapy is effective. The set of reference levels of expression may be in the form of a database. The database may be included in a computer-readable medium. The database may be in communications with a microprocessor and microprocessor instructions for providing a user interface to receive expression level data of a subject and to compare the expression level data with the database. In a particular embodiment, the method comprises (i) obtaining a patient sample from a caregiver; (ii) identifying expression levels of a plurality of

genes characteristic of R.A. from the patient sample; (iii) determining whether the levels of expression of the genes in the patient sample are more similar to those of a subject having R.A. or to those of a subject who does not have R.A.; and (iv) transmitting the results to the caregiver. The results may be transmitted across a network.

5 In yet another embodiment, the invention provides methods for identifying a compound for treating R.A. The method comprises, e.g., (i) providing levels of expression of a plurality of genes characteristic of R.A. in a cell characteristic of R.A. incubated with a test compound; (ii) providing levels of expression of a normal counterpart cell; and (iii) comparing the two levels of expression, wherein similar levels of expression in the two cells indicates that the compound is
10 likely to be effective for treating R.A.

15 Other methods provided by the invention include methods for selecting a therapy for a patient having R.A. For example, the method may comprise (i) providing at least one query value corresponding to the level of expression of at least one gene characteristic of R.A. from a patient having R.A.; (ii) providing a plurality of sets of reference values corresponding to levels of expression of at least one gene characteristic of R.A., each reference value being associated with a therapy; and (iii) selecting the reference values most similar to the query values, to thereby select a therapy for said patient. Selecting may further include weighing a comparison value for the reference values using a weight value associated with each reference values. The method may further comprise administering the therapy to the patient. The query values and the
20 sets of reference values may be expression profiles. Another exemplary method comprises (i) providing a plurality of reference expression profiles, each associated with a therapy; (ii) providing a labeled target nucleic acid sample prepared from RNA of a diseased cell of the patient; (iii) contacting the labeled target nucleic acid sample with an array comprising probes corresponding to genes which are up- or down-regulated in R.A. to obtain an expression profile
25 of the patient; and selecting the reference profile most similar to the expression profile of the patient, to thereby select a therapy for the patient.

The invention also provides therapeutic methods for treating R.A., including methods which normalize the expression level of one or more genes characteristic of R.A. in a subject diagnosed with R.A. "Normalization" of the level of expression of a gene refers to a change in
30 the expression level of the gene such that its level of its expression resembles more that of a non-diseased (i.e., normal) cell than that of a diseased cell. Such methods may include administering

to a subject having R.A. a pharmaceutically efficient amount of an agonist or antagonist of one or more genes characteristic of R.A.

Also within the scope of the invention are diagnostic or drug discovery kits, comprising one or more computer-readable media, compositions and/or solid surfaces described herein, and optionally instructions for use.

Detailed Description of the Invention

The invention is based at least in part on the discovery of gene expression profiles of cells of subjects having R.A. As described in the Examples and in Tables 1-5, cells from R.A. subjects have genes which are expressed at higher levels (i.e., which are up-regulated) and genes which are expressed at lower levels (i.e., which are down-regulated) relative to cells of the same type in subjects which do not have any symptoms of R.A. In particular, as described in the Examples, it has been shown that genes SOCS3 (CISH3); RAGE (AGER); LST-1 (LY117); SAA 1-3; HMG-1; S100 A8, A9, and A12; SLPI; GILZ; PTPN-18; GADD-45A and B; Legumain (PRSC1); FST1; Lcn2; GPI; SpiL; and TSG-6 are expressed at higher levels in the diseased cells relative to the corresponding normal cells. Other genes, e.g., CMAK2B; PLA2G2A; GBAS and SOX15, are down-regulated in the diseased cells relative to the corresponding normal cells.

1. Definitions:

As used herein, the following terms and phrases shall have the meanings set forth below. Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood to one of ordinary skill in the art to which this invention belongs.

The singular forms “a,” “an,” and “the” include plural reference unless the context clearly dictates otherwise.

The phrase “a corresponding normal cell of” or “normal cell corresponding to” or “normal counterpart cell of” a diseased cell refers to a normal cell of the same type as that of the diseased cell. For example, a corresponding normal PBMC of a subject having R.A. is a PBMC of a subject not having R.A.

An “address” on an array, e.g., a microarray, refers to a location at which an element, e.g., an oligonucleotide, is attached to the solid surface of the array.

The term "agonist," as used herein, is meant to refer to an agent that mimics or up-regulates (e.g., potentiates or supplements) the bioactivity of a protein. An agonist can be a wild-type protein or derivative thereof having at least one bioactivity of the wild-type protein. An agonist can also be a compound that upregulates expression of a gene or which increases at least one bioactivity of a protein. An agonist can also be a compound which increases the interaction of a polypeptide with another molecule, e.g., a target peptide or nucleic acid.

"Amplification," as used herein, relates to the production of additional copies of a nucleic acid sequence. Amplification is generally carried out using polymerase chain reaction (PCR) technologies well known in the art. (Dieffenbach, C. W. and G. S. Dveksler (1995) PCR Primer, a Laboratory Manual, Cold Spring Harbor Press, Plainview, N.Y.)

"Antagonist" as used herein is meant to refer to an agent that downregulates (e.g., suppresses or inhibits) at least one bioactivity of a protein. An antagonist can be a compound which inhibits or decreases the interaction between a protein and another molecule, e.g., a target peptide or enzyme substrate. An antagonist can also be a compound that downregulates expression of a gene or which reduces the amount of expressed protein present.

The term "antibody" as used herein is intended to include whole antibodies, e.g., of any isotype (IgG, IgA, IgM, IgE, etc), and includes fragments thereof which are also specifically reactive with a vertebrate, e.g., mammalian, protein. Antibodies can be fragmented using conventional techniques and the fragments screened for utility in the same manner as described above for whole antibodies. Thus, the term includes segments of proteolytically-cleaved or recombinantly-prepared portions of an antibody molecule that are capable of selectively reacting with a certain protein. Nonlimiting examples of such proteolytic and/or recombinant fragments include Fab, F(ab')₂, Fab', Fv, and single chain antibodies (scFv) containing a V[L] and/or V[H] domain joined by a peptide linker. The scFv's may be covalently or non-covalently linked to form antibodies having two or more binding sites. The subject invention includes polyclonal, monoclonal, or other purified preparations of antibodies and recombinant antibodies.

By "array" or "matrix" is meant an arrangement of addressable locations or "addresses" on a device. The locations can be arranged in two dimensional arrays, three dimensional arrays, or other matrix formats. The number of locations can range from several to at least hundreds of thousands. Most importantly, each location represents a totally independent reaction site. A "nucleic acid array" refers to an array containing nucleic acid probes, such as oligonucleotides or

larger portions of genes. The nucleic acid on the array is preferably single stranded. Arrays wherein the probes are oligonucleotides are referred to as "oligonucleotide arrays" or "oligonucleotide chips." A "microarray," also referred to herein as a "biochip" or "biological chip" is an array of regions having a density of discrete regions of at least about 100/cm², and preferably at least about 1000/cm². The regions in a microarray have typical dimensions, e.g., diameters, in the range of between about 10-250 μm, and are separated from other regions in the array by about the same distance.

The term "biological sample", as used herein, refers to a sample obtained from a subject, e.g., a human or from components (e.g., tissues) of a subject. The sample may be of any biological tissue or fluid. Frequently the sample will be a "clinical sample" which is a sample derived from a patient. Such samples include, but are not limited to, sputum, blood, blood cells (e.g., white cells), tissue or fine needle biopsy samples, urine, peritoneal fluid, and pleural fluid, or cells therefrom. Biological samples may also include sections of tissues such as frozen sections taken for histological purposes. A preferred biological sample is a PBMC sample or a sample from a joint, e.g., synovial fluid or synovial tissue.

The term "biomarker" of a disease refers to a gene which is up- or down-regulated in a diseased cell of a subject having R.A. relative to a counterpart normal cell, which gene is sufficiently specific to the diseased cell that it can be used, optionally with other genes, to identify or detect the disease. Generally, a biomarker is a gene that is characteristic of the disease.

A nucleotide sequence is "complementary" to another nucleotide sequence if each of the bases of the two sequences match, i.e., are capable of forming Watson-Crick base pairs. The term "complementary strand" is used herein interchangeably with the term "complement." The complement of a nucleic acid strand can be the complement of a coding strand or the complement of a non-coding strand.

A "computer readable medium" is any medium that can be used to store data which can be accessed by a computer. Exemplary media include: magnetic storage media, such as a diskettes, hard drives, and magnetic tape; optical storage media such as CD-ROMs; electrical storage media such as RAM and ROM; and hybrids of these media, such as magnetic/optical storage medium.

A “cell characteristic of R.A.” refers to a cell present in subjects having R.A., which cell is a modified form of a normal cell and is not present in a subject not having R.A., or which cell is present in significantly higher or lower numbers in subjects having R.A. relative to subjects not having R.A. A “modified form of a normal cell” can be a form of the normal cell in which the expression of at least one gene is higher or lower (e.g., by 50%, 2 fold, 5 fold, or over 10 fold) relative to the normal cell. A cell characteristic of R.A. is also referred to herein as a “diseased cell of R.A.” Exemplary diseased cells of R.A. include PBMCs, e.g., monocytes and macrophages, and inflammatory cells present in joints of patients, in particular, in synovial fluid and synovium. Inflammatory cells can be lymphocytes, e.g., T lymphocytes, B lymphocytes, monocytes and macrophages. Other diseased cells of R.A. include neutrophils, fibroblasts, endothelial cells, osteoclasts, osteoblasts, osteocytes, chondrocytes, and cells present in cartilage.

A “cell corresponding to a cell characteristic of R.A.” refers to a cell which has essentially the same phenotype as that of a cell characteristic of R.A. For example, a cell corresponding to a PBMC or a subject having R.A. is a PBMC of a subject who does not have R.A.

A “cell sample characteristic of R.A.” or a “tissue sample characteristic of R.A.” refers to a sample of cells, such as a tissue, that contains at least one cell characteristic of R.A. Such a sample may be a sample of blood, PBMCs, synovial fluid, synovium, cartilage or bone.

The term “derivative” refers to the chemical modification of a compound, e.g., a polypeptide, or a polynucleotide. Chemical modifications of a polynucleotide can include, for example, replacement of hydrogen by an alkyl, acyl, or amino group. A derivative polynucleotide encodes a polypeptide which retains at least one biological or immunological function of the natural molecule. A derivative polypeptide can be one modified by glycosylation, pegylation, or any similar process that retains at least one biological or immunological function of the polypeptide from which it was derived.

A “detection agent of a gene” refers to an agent that can be used to specifically detect a gene or other biological molecule relating to it, e.g., RNA transcribed from the gene and polypeptides encoded by the gene. Exemplary detection agents are nucleic acid probes which hybridize to nucleic acids corresponding to the gene and antibodies.

The term “equivalent” is understood to include nucleotide sequences encoding functionally equivalent polypeptides. Equivalent nucleotide sequences will include sequences

that differ by one or more nucleotide substitutions, additions or deletions, such as allelic variants; and will, therefore, include sequences that differ from the nucleotide sequence of the nucleic acids referred to in Any of Tables 1-5 due to the degeneracy of the genetic code.

5 The term “essentially all the genes of any of Tables 1-5” refers to at least 90%, preferably at least 95% and most preferably at least 98% of the genes of any of Tables 1-5.

10 The term “expression profile,” which is used interchangeably herein with “gene expression profile” and “finger print” refers to a set of values representing the activity of about 10 or more genes. An expression profile preferably comprises values representing expression levels of at least about 20 genes, preferably at least about 30, 50, 100, 200 or more genes. An expression profile can be a set of values obtained from one or more cells or from a tissue sample, e.g., a clinical sample. An expression profile of a cell characteristic of R.A. may refer to a set of values representing mRNA levels of about 10 or more genes in a cell characteristic of R.A. An “expression profile of R.A.” refers to an expression profile of a cell characteristic of R.A. Thus, since there are different cells characteristic of R.A., there may be different expression profiles of R.A.

15 “Genes which are up- or down-regulated in R.A.” refers to genes which are up- or down-regulated in cells characteristic of R.A. relative to normal counterpart cells.

20 “Genes characteristic of R.A.” refers to genes which are up- or down-regulated by a significant factor, e.g., at least about 1.1 fold, 1.25 fold, 1.5 fold, 2 fold, 5 fold, 10 fold or more in at least about 50%, preferably 60%, 70%, 80%, or 90% of subjects having R.A., as determined, e.g., by methods described herein. Preferred genes characteristic of R.A. are those described in Tables 1-5. Even more preferred genes are those which are highlighted or marked with a star in the Tables, those which encode kinases or phosphatases and those which are located on human chromosome 6, preferably at 6p21.3.

25 “Hybridization” refers to any process by which a strand of nucleic acid binds with a complementary strand through base pairing. Two single-stranded nucleic acids “hybridize” when they form a double-stranded duplex. The region of double-strandedness can include the full-length of one or both of the single-stranded nucleic acids, or all of one single stranded nucleic acid and a subsequence of the other single stranded nucleic acid, or the region of double-strandedness can include a subsequence of each nucleic acid. Hybridization also includes the formation of duplexes which contain certain mismatches, provided that the two strands are still

30

forming a double stranded helix. "Stringent hybridization conditions" refers to hybridization conditions resulting in essentially specific hybridization.

The term "isolated" as used herein with respect to nucleic acids, such as DNA or RNA, refers to molecules separated from other DNAs, or RNAs, respectively, that are present in the natural source of the macromolecule. The term isolated as used herein also refers to a nucleic acid or peptide that is substantially free of cellular material, viral material, or culture medium when produced by recombinant DNA techniques, or chemical precursors or other chemicals when chemically synthesized. Moreover, an "isolated nucleic acid" is meant to include nucleic acid fragments which are not naturally occurring as fragments and would not be found in the natural state. The term "isolated" is also used herein to refer to polypeptides which are isolated from other cellular proteins and is meant to encompass both purified and recombinant polypeptides.

As used herein, the terms "label" and "detectable label" refer to a molecule capable of detection, including, but not limited to, radioactive isotopes, fluorophores, chemiluminescent moieties, enzymes, enzyme substrates, enzyme cofactors, enzyme inhibitors, dyes, metal ions, ligands (e.g., biotin or haptens) and the like. The term "fluorescer" refers to a substance or a portion thereof which is capable of exhibiting fluorescence in the detectable range. Particular examples of labels which may be used under the invention include fluorescein, rhodamine, dansyl, umbelliferone, Texas red, luminol, NADPH, alpha - beta -galactosidase and horseradish peroxidase.

The "level of expression of a gene in a cell" refers to the activity of a gene in the cell, which can be indicated by the level of mRNA, as well as pre-mRNA nascent transcript(s), transcript processing intermediates, mature mRNA(s) and degradation products, encoded by the gene in the cell.

The phrase "normalizing expression of a gene" in a diseased cell refers to an action to compensate for the altered expression of the gene in the diseased cell, so that it is essentially expressed at the same level as in the corresponding non diseased cell. For example, where the gene is over-expressed in the diseased cell, normalization of its expression in the diseased cell refers to treating the diseased cell in such a way that its expression becomes essentially the same as the expression in the counterpart normal cell. "Normalization" preferably brings the level of expression to within approximately a 50% difference in expression, more preferably to within

approximately a 25%, and even more preferably 10% difference in expression. The required level of closeness in expression will depend on the particular gene, and can be determined as described herein. The phrase "normalizing gene expression in a diseased cell" refers to an action to normalize the expression of essentially all genes in the diseased cell.

As used herein, the term "nucleic acid" refers to polynucleotides such as deoxyribonucleic acid (DNA), and, where appropriate, ribonucleic acid (RNA). The term should also be understood to include, as equivalents, analogs of either RNA or DNA made from nucleotide analogs, and, as applicable to the embodiment being described, single (sense or antisense) and double-stranded polynucleotides. ESTs, chromosomes, cDNAs, mRNAs, and rRNAs are representative examples of molecules that may be referred to as nucleic acids.

The phrase "nucleic acid corresponding to a gene" refers to a nucleic acid that can be used for detecting the gene, e.g., a nucleic acid which is capable of hybridizing specifically to the gene.

The phrase "nucleic acid sample derived from RNA" refers to one or more nucleic acid molecule, e.g., RNA or DNA, that was synthesized from the RNA, and includes DNA resulting from methods using PCR, e.g., RT-PCR.

The term "percent identical" refers to sequence identity between two amino acid sequences or between two nucleotide sequences. Identity can each be determined by comparing a position in each sequence which may be aligned for purposes of comparison. When an equivalent position in the compared sequences is occupied by the same base or amino acid, then the molecules are identical at that position; when the equivalent site occupied by the same or a similar amino acid residue (e.g., similar in steric and/or electronic nature), then the molecules can be referred to as homologous (similar) at that position. Expression as a percentage of homology, similarity, or identity refers to a function of the number of identical or similar amino acids at positions shared by the compared sequences. Various alignment algorithms and/or programs may be used, including FASTA, BLAST, or ENTREZ. FASTA and BLAST are available as a part of the GCG sequence analysis package (University of Wisconsin, Madison, Wis.), and can be used with, e.g., default settings. ENTREZ is available through the National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, Md. In one embodiment, the percent identity of two sequences can be determined by the GCG program with a gap weight of 1, e.g., each amino acid gap is weighted as

if it were a single amino acid or nucleotide mismatch between the two sequences. Other techniques for alignment are described in Methods in Enzymology, vol. 266: Computer Methods for Macromolecular Sequence Analysis (1996), ed. Doolittle, Academic Press, Inc., a division of Harcourt Brace & Co., San Diego, California, USA. Preferably, an alignment program that permits gaps in the sequence is utilized to align the sequences. The Smith-Waterman is one type of algorithm that permits gaps in sequence alignments. See Meth. Mol. Biol. 70: 173-187 (1997). Also, the GAP program using the Needleman and Wunsch alignment method can be utilized to align sequences. An alternative search strategy uses MPSRCH software, which runs on a MASPAR computer. MPSRCH uses a Smith-Waterman algorithm to score sequences on a massively parallel computer. This approach improves ability to pick up distantly related matches, and is especially tolerant of small gaps and nucleotide sequence errors. Nucleic acid-encoded amino acid sequences can be used to search both protein and DNA databases. Databases with individual sequences are described in Methods in Enzymology, ed. Doolittle, *supra*. Databases include Genbank, EMBL, and DNA Database of Japan (DDBJ).

“Perfectly matched” in reference to a duplex means that the poly- or oligonucleotide strands making up the duplex form a double stranded structure with one other such that every nucleotide in each strand undergoes Watson-Crick basepairing with a nucleotide in the other strand. The term also comprehends the pairing of nucleoside analogs, such as deoxyinosine, nucleosides with 2-aminopurine bases, and the like, that may be employed. A mismatch in a duplex between a target polynucleotide and an oligonucleotide or oligonucleotide means that a pair of nucleotides in the duplex fails to undergo Watson-Crick bonding. In reference to a triplex, the term means that the triplex consists of a perfectly matched duplex and a third strand in which every nucleotide undergoes Hoogsteen or reverse Hoogsteen association with a basepair of the perfectly matched duplex.

A “plurality” refers to two or more.

As used herein, a nucleic acid or other molecule attached to an array, is referred to as a “probe” or “capture probe.” When an array contains several probes corresponding to one gene, these probes are referred to as “gene-probe set.” A gene-probe set can consist of, e.g., 2 to 10 probes, preferably from 2 to 5 probes and most preferably about 5 probes.

The “profile” of a cell’s biological state refers to the levels of various constituents of a cell that are known to change in response to drug treatments and other perturbations of the cell’s

biological state. Constituents of a cell include levels of RNA, levels of protein abundances, or protein activity levels.

The term “protein” is used interchangeably herein with the terms “peptide” and “polypeptide.”

5 “Rheumatoid arthritis” or “R.A.” refers to a systemic chronic inflammatory disease involving primarily the joints of the extremities. It is characterized by destruction of the joint cartilage and inflammation of the synovium, with a morphologic picture suggestive of a local immune response. CD4⁺ T cells, activated B lymphocytes and plasma cells are found in the inflamed synovium, and in severe cases, well formed lymphoid follicles with germinal centers
10 may be present. The synovial fluid and serum contain rheumatoid factors, i.e., complexes containing auto-antibodies, and may cytokines, e.g., interleukin-1 (IL-1), tumor necrosis factor (TNF) and interferon gamma (IFN- γ). T cells expressing the $\gamma\delta$ antigen receptor are also present in the synovial fluid of patients. R.A. is described, e.g., in Cecil Essentials of Medicine, Third Edition, Andreoli et al., W.B. Saunders Company (1993) at pages 564 to 568. This reference describes in particular symptoms that are the basis of a diagnosis of R.A. This reference also describes the different stages of the disease. Briefly, the first stage is characterized by presentation of antigen to T cells and is not associated with any symptoms. The second stage is characterized by T- and B-cell proliferation and angiogenesis in synovial membrane. The symptoms of the second stage are malaise, mild joint stiffness and swelling. The third stage is characterized by accumulation of neutrophils in synovial fluid; synovial cell proliferation without polarization or invasion or cartilage. The symptoms in this stage are joint pain and swelling; morning stiffness, malaise and weakness. The fourth stage is characterized by polarization of synovitis into a centripetally invasive pannus; activation of chondrocytes; initiation of enzyme (proteinase) degradation of cartilage. The symptoms in this stage are the
25 same as those associated with stage three. The fifth stage is characterized by erosion of subchondral bone; invasion of cartilage by pannus; chondrocyte proliferation; and stretched ligaments around joints. The symptoms in this stage are the same as those associated with stage 3, and in addition, loss of function and early deformity (e.g., ulnar deviation at metacarpophalangeal joint). Therapeutics used for treating R.A. include aspirin or other non-steroidal anti-inflammatory drug (NSAID); immunosuppressive agents, e.g., azathioprine, cyclophosphamide, chlorambucil and methotrexate; corticosteroids; gold salts; penicillamine;
30

InfliximabTM (anti-TNF antibody); EtanerceptTM or EnbrelTM (a soluble TNF receptor); LeflunomideTM; Anakinra (IL-1 antagonist); and KinaretTM (IL-1 antagonist).

A "similarity" between the level of expression of a gene in two cells or tissues refers to a difference in expression levels of a factor of at least about 10% (i.e., 1.1 fold), 25% (i.e., 1.25 fold), 50% (i.e., 1.5 fold), 75% (i.e., 1.75 fold), 90% (i.e., 1.9 fold), 2 fold, 2.5 fold, 3 fold, 5 fold, 10 fold, 50 fold, or 100 fold. Expression levels can be raw data or they can be averaged or normalized data, e.g., normalized relative to normal controls.

An expression profile in one cell or tissue is "similar" to an expression profile in another cell or tissue when the level of expression of the genes in the two expression profiles are sufficiently similar that the similarity is indicative of a common characteristic, e.g., being of the same cell type, or being characteristic of R.A. "Similarity" between an expression profile of a cell or tissue, e.g., of a subject, and a set of data representing an expression profile characteristic of a disease can be based on the presence or absence in the cell or tissue of certain RNAs and/or certain levels of certain RNAs of genes having a high probability of being associated with the disease. A high probability of being associated with a disease can be, e.g., the presence of RNA or of certain levels of RNA of particular genes which are over-expressed or under-expressed, in at least about 50%, 60%, 70%, 80%, 90%, or 100% of patients having the disease. A similarity with an expression profile of a patient can also be based on higher or lower expression levels of a factor of at least about 10%, 25%, 50%, 75%, 1.5 fold, 2 fold, 2.5 fold, 3 fold, 5 fold, 10 fold, 50 fold, 100 fold of at least about 50%, 60%, 70%, 80%, 90%, or 100% of genes, or at least about 10, 50, 100, 200, 300 genes, which are up- or down-regulated in at least about 50%, 60%, 70%, 80%, 90%, or 100% of patients. For example, the expression profile of PBMCs of a subject is similar to a reference expression profile for an R.A. patient, e.g., determined herein, if at least about 50 genes which are over-expressed or repressed (i.e., under-expressed), e.g., at least about 1.1 fold, in at least about 60% of the patients studied are over-expressed or repressed, e.g., about at least 1.1 fold, in the expression profile of the subject. A similarity in expression profiles may also include similar expression levels of genes which are not up- or down-regulated in R.A.

"Small molecule" as used herein, is meant to refer to a composition, which has a molecular weight of less than about 5 kD and most preferably less than about 4 kD. Small molecules can be nucleic acids, peptides, polypeptides, peptidomimetics, carbohydrates, lipids or other organic (carbon-containing) or inorganic molecules. Many pharmaceutical companies

have extensive libraries of chemical and/or biological mixtures, often fungal, bacterial, or algal extracts, which can be screened with any of the assays of the invention to identify compounds that modulate a bioactivity.

The term “specific hybridization” of a probe to a target site of a template nucleic acid refers to hybridization of the probe predominantly to the target, such that the hybridization signal can be clearly interpreted. As further described herein, such conditions resulting in specific hybridization vary depending on the length of the region of homology, the GC content of the region, the melting temperature “T_m” of the hybrid. Hybridization conditions will thus vary in the salt content, acidity, and temperature of the hybridization solution and the washes.

A “subject” can be a mammal, e.g., a human, primate, ovine, bovine, porcine, equine, feline, and canine.

The term “treating” a disease in a subject or “treating” a subject having a disease refers to providing the subject with a pharmaceutical treatment, e.g., the administration of a drug, such that at least one symptom of the disease is decreased. Treating a disease can be preventing the disease, improving the disease or curing the disease. Treatment of R.A. includes inhibition of erosion, e.g., cartilage or bone erosion, and/or inhibition of inflammation.

The phrase “value representing the level of expression of a gene” refers to a raw number which reflects the mRNA level of a particular gene in a cell or biological sample, e.g., obtained from analytical tools for measuring RNA levels.

A “variant” of a polypeptide refers to a polypeptide having the amino acid sequence of the polypeptide, in which one or more amino acid residues are altered. The variant may have “conservative” changes, wherein a substituted amino acid has similar structural or chemical properties (e.g., replacement of leucine with isoleucine). More rarely, a variant may have “non-conservative” changes (e.g., replacement of glycine with tryptophan). Analogous minor variations may also include amino acid deletions or insertions, or both. Guidance in determining which amino acid residues may be substituted, inserted, or deleted without abolishing biological or immunological activity may be found using computer programs well known in the art, for example, LASERGENE software (DNASTAR). The term “variant,” when used in the context of a polynucleotide sequence, encompasses a polynucleotide sequence related to that of a gene of interest or the coding sequence thereof. This definition may also include, for example, “allelic,” “splice,” “species,” or “polymorphic” variants. A splice variant may have significant identity to

a reference molecule, but will generally have a greater or lesser number of polynucleotides due to alternate splicing of exons during mRNA processing. The corresponding polypeptide may possess additional functional domains or an absence of domains. Species variants are polynucleotide sequences that vary from one species to another. The resulting polypeptides generally will have significant amino acid identity relative to each other. A polymorphic variant is a variation in the polynucleotide sequence of a particular gene between individuals of a given species. Polymorphic variants also may encompass "single nucleotide polymorphisms" (SNPs) in which the polynucleotide sequence varies by one base. The presence of SNPs may be indicative of, for example, a certain population, a disease state, or a propensity for a disease state.

2. R.A. diagnostic and prognostic methods of use

The invention provides gene expression profiles of R.A. As further described herein, the gene expression profiles of the diseased cells of subjects having R.A., indicates that genes certain genes, e.g., SOCS3 (CISH3); RAGE (AGER); LST-1 (LY117); SAA 1-3; HMG-1; S100 A8, A9, and A12; SLPI; GILZ; PTPN-18; GADD-45A and B; Legumain (PRSC1); FST1; Lcn2; GPI; SpiL; and TSG-6, are significantly up-regulated in these cells relative to their normal counterparts. The expression data also show that certain genes, e.g., CAMK2B, PLA2G2A, GBAS and SOX15, are significantly down-regulated in the diseased cells relative to their normal counterpart cells. Other preferred genes include those that are highlighted or marked with a star in Tables 1-5. Yet other genes of particular interest are those that have a fold induction indicated as "#DIV/0!" in the Tables; those that encode kinases and phosphatases; those that are localized to human chromosome 6p21.3; and those which are highlighted or marked with a star in the Tables. Accordingly, the expression profile can be used diagnostically and prognostically for R.A. Exemplary diagnostic tools and assays are set forth below, under (i) to (vi), followed by exemplary methods for conducting these assays.

Preferred methods of the invention involve measuring the level of expression of one or more genes that are up- or down-regulated in R.A. in a cell of a patient, and comparing these levels of expression to the level of expression of the genes in other samples, which levels of expression may be present in a computer readable medium and analyzed with a computer.

(i) In one embodiment, the invention provides a method for determining whether a subject has or is likely to develop R.A., comprising determining the level of expression of one or more genes which are up- or down-regulated in R.A. in a cell of the subject and comparing these levels of expression with the levels of expression of the genes in a diseased cell of a subject known to have R.A. A similar level of expression of the genes in the two cells is indicative that the subject has or is likely to develop R.A. or at least a symptom thereof. In a preferred embodiment, the cell of the subject is essentially of the same type as that which is diseased in R.A.

(ii) In another embodiment the expression profile data of the invention can be used to confirm that a subject has R.A., and in particular, that the subject does not have a disease that is merely related R.A. This can be important, in particular, in designing an optimal therapeutic regimen for the subject. It has been described in the art that expression profiles can be used to distinguish one type of disease from a similar disease. For example, two subtypes of non-Hodgkin's lymphomas, one of which responds to current therapeutic methods and the other one which does not, could be differentiated by investigating 17,856 genes in specimens of patients suffering from diffuse large B-cell lymphoma (Alizadeh et al. (2000) Nature 405:503). Similarly, subtypes of cutaneous melanoma were predicted based on profiling 8150 genes (Bittner et al. (2000) Nature 406:536). In this case, features of the highly aggressive metastatic melanomas could be recognized. Numerous other studies comparing expression profiles of cancer cells and normal cells have been described, including studies describing expression profiles distinguishing between highly and less metastatic cancers and studies describing new subtypes of diseases, e.g., new tumor types (*see, e.g.*, Perou et al. (1999) PNAS 96: 9212; Perou et al. (2000) Nature 606:747; Clark et al. (2000) Nature 406:532; Alon et al. (1999) PNAS 96:6745; Golub et al. (1999) Science 286:531).

Accordingly, the expression profiles of the invention allow the distinction of R.A. from related diseases. In a preferred embodiment, the level of expression of one or more genes which are up- or down-regulated in R.A. is determined in a cell of the subject, preferably a cell which corresponds to a diseased cell in R.A. A level of expression of one or more genes that is more similar to that in a cell characteristic of R.A. than to that of cells of related diseases indicates that the subject has R.A., rather than a disease related to R.A.

Prior to using this method for determining whether the subject has R.A. or a related disease, it may be necessary to first determine the expression profile of cells of diseases that are

similar to R.A. This can be undertaken using the same microarray as the one that was used to identify the genes characteristic of R.A., and according to methods further described herein.

(iii) In yet another embodiment, the invention provides methods for determining the stage of R.A. in the subject. In one embodiment, the level of expression of one or more genes that are up- or down-regulated in R.A., in particular, whose level of expression varies with the stage of the disease is determined in a cell of a subject. A level of expression of one or more genes that is more similar to that of one stage of the disease (stage "a") relative to that in other stages of the disease indicates that the disease of the subject is in stage a.

This assay may require the preliminary determination of expression profiles in different stages of R.A. Such expression data can be obtained by, e.g., using microarrays with target nucleic acids made from RNA of patients at different stages of the disease.

(iv) The method can also be used to determine the efficacy of a therapy in a subject. Accordingly, in one embodiment, the level of expression of one or more genes which are up- or down-regulated in R.A. is determined in a subject before the treatment and one or more times during the treatment. For example, a sample of RNA can be obtained from the subject before the beginning of the therapy and every 12, 24 or 72 hours during the therapy. Samples can also be analyzed once a week or once a month. Changes in expression levels of the genes over time and relative to diseased cells and normal cells will indicate whether the therapy is effective. For example, expression levels that are more similar to those in normal cells or in less advanced stages of the disease relative to the stage the subject was in, indicates that the therapy is effective.

(v) In yet another embodiment, the invention provides a method for determining the likelihood of success of a particular therapy in a subject having R.A. In one embodiment, a subject is started on a particular therapy, and the effectiveness of the therapy is determined, e.g., by determining the level of expression of one or more genes characteristic of R.A. in a cell of the subject. A normalization of the level of expression of these genes, i.e., a change in the expression level of the genes such that their level of expression resembles more that of a non diseased cell, indicates that the treatment should be effective in the subject. On the other hand, the absence of normalization of the level of expression of the genes characteristic of R.A. indicates that the treatment is not likely to be effective in the subject. This method may be able to predict that a treatment is effective before any alleviation of symptoms becomes apparent.

Prediction of the outcome of a treatment of R.A. in a subject can also be undertaken *in vitro*. In one embodiment, cells are obtained from a subject to be evaluated for responsiveness to the treatment, and incubated *in vitro* with the therapeutic drug or metabolized form thereof. The level of expression of one or more genes which are up- or down-regulated in R.A. is measured in the cells and these values are compared to the level of expression of these one or more genes in a cell which is a normal counterpart cell of a cell characteristic of R.A. The level of expression can also be compared to that in other diseased cells. A level of expression of one or more genes in the cells of the subject after incubation with the drug that is similar to their level of expression in a normal cell and different from that in a diseased cell is indicative that it is likely that the subject will respond positively to a treatment with the drug. On the contrary, levels of expressions that are more similar to levels of expression in a diseased cell than that in a normal cell is indicative that it is likely that the subject will not respond positively to a treatment with the drug.

Since it is possible that a drug for treating R.A. does not act directly on the diseased cells, but is, e.g., metabolized, or acts on another cell which then secretes a factor that will effect the diseased cells, the above assay can also be conducted in a tissue sample of a subject, which contains cells other than the diseased cells. For example, a tissue sample comprising diseased cells is obtained from a subject; the tissue sample is incubated with the potential drug; optionally one or more diseased cells are isolated from the tissue sample, e.g., by microdissection or Laser Capture Microdissection (LCM, *see infra*); and the expression level of one or more genes characteristic of R.A. is examined.

(vi) The invention also provides methods for selecting a particular therapy for an R.A. patient from a selection of several different therapies. Certain subjects having R.A. may respond better to one type of therapy than to another type of therapy. In a preferred embodiment, the method comprises comparing the expression level of at least one gene that is up- or down-regulated in R.A. in the patient with that in cells of R.A. subjects that were treated *in vitro* or *in vivo* with one of several therapeutic drugs, which subjects are responders or non responders to one of the therapeutic drugs, and identifying the cell which has the most similar level of expression of that in the patient, to thereby identify a therapy for the patient. The method may further comprise administering the therapy to the subject.

A person of skill in the art will recognize that in certain diagnostic and prognostic assays, it will be sufficient to assess the level of expression of a single gene that is up- or down-regulated in R.A., and that in others, the expression of a plurality, e.g., two or more genes, is preferred. In certain embodiments, it is preferable to assess the expression of at least about 10%, or at least about 20%, 30%, 50%, 70%, 90% or 95% of the genes listed in one or more of Tables 1-5 or of the genes characteristic of R.A.

A person of skill in the art will also recognize that expression levels can be measured in a single cell or in a plurality of cells, e.g., two or more cells. In one embodiment, the method comprises determining expression levels in a cell or tissue sample, e.g., a blood sample, a PBMC sample, a synovial fluid sample or a synovium sample.

Set forth below are exemplary methods which can be used to determine the level of expression of one or more genes. In a preferred embodiment for determining the level of expression of a plurality of genes, arrays, e.g., microarrays, can be used.

2.1. Use of arrays for determining the level of expression of genes

Generally, determining expression profiles with arrays involves the following steps: (a) obtaining a mRNA sample from a subject and preparing labeled nucleic acids therefrom (the "target nucleic acids" or "targets"); (b) contacting the target nucleic acids with the array under conditions sufficient for target nucleic acids to bind with corresponding probes on the array, e.g. by hybridization or specific binding; (c) optionally removing unbound targets from the array; (d) detecting bound targets, and (e) analyzing the results. As used herein, "nucleic acid probes" or "probes" are nucleic acids attached to the array, whereas "target nucleic acids" are nucleic acids that are hybridized to the array. Each of these steps is described in more detail below.

(i) *Obtaining a mRNA sample of a subject*

In one embodiment, one or more cells from the subject to be tested are obtained and RNA is isolated from the cells. In a preferred embodiment, PBMCs, synovial fluid, synovium or cartilage are obtained from the subject according to methods known in the art. Examples of such methods are set forth in the Examples and is discussed by Kim, C. H. et al. (J. Virol. 66:3879-3882 (1992)); Biswas, B. et al. (Annals NY Acad. Sci. 590:582-583 (1990)); Biswas, B. et al. (J. Clin. Microbiol. 29:2228-2233 (1991)). When obtaining the cells, it is preferable to obtain a sample containing predominantly cells of the desired type, e.g., a sample of cells in which at

least about 50%, preferably at least about 60%, even more preferably at least about 70%, 80% and even more preferably, at least about 90% of the cells are of the desired type. A higher percentage of cells of the desired type is preferable, since such a sample is more likely to provide clear gene expression data.

5 It is also possible to obtain a cell sample from a subject, and then to enrich it for a desired cell type. For example, PBMCs can be isolated from blood as described herein. Counter-flow centrifugation (elutriation) can also be used to enrich for various cell types, such as T cells, B cells and monocytes, from PBMCs. Cells can also be isolated from other cells using a variety of techniques, such as isolation with an antibody binding to an epitope on the cell surface of the
10 desired cell type. Another method that can be used includes negative selection using antibodies to cell surface markers to selectively enrich for a specific cell type without activating the cell by receptor engagement. Where the desired cells are in a solid tissue, particular cells can be dissected out, e.g., by microdissection. Exemplary cells that one may want to enrich for include monocytes, macrophages, T and B cells, osteocytes, osteoblasts, osteoclasts, chondrocytes, fibroblasts, neutrophils, endothelial cells and other cartilage cells.

In one embodiment, RNA is obtained from a single cell. For example, a cell can be isolated from a tissue sample by laser capture microdissection (LCM). Using this technique, a cell can be isolated from a tissue section, including a stained tissue section, thereby assuring that the desired cell is isolated (*see, e.g.*, Bonner et al. (1997) Science 278: 1481; Emmert-Buck et al.
15 (1996) Science 274:998; Fend et al. (1999) Am. J. Path. 154: 61 and Murakami et al. (2000) Kidney Int. 58:1346). For example, Murakami et al., *supra*, describe isolation of a cell from a previously immunostained tissue section.

It is also possible to obtain cells from a subject and culture the cells *in vitro*, such as to obtain a larger population of cells from which RNA can be extracted. Methods for establishing
20 cultures of non-transformed cells, i.e., primary cell cultures, are known in the art.

When isolating RNA from tissue samples or cells from individuals, it may be important to prevent any further changes in gene expression after the tissue or cells has been removed from the subject. Changes in expression levels are known to change rapidly following perturbations, e.g., heat shock or activation with lipopolysaccharide (LPS) or other reagents. In addition, the
25 RNA in the tissue and cells may quickly become degraded. Accordingly, in a preferred embodiment, the tissue or cells obtained from a subject is snap frozen as soon as possible.

RNA can be extracted from the tissue sample by a variety of methods, e.g., those described in the Examples or guanidium thiocyanate lysis followed by CsCl centrifugation (Chirgwin et al., 1979, Biochemistry 18:5294-5299). RNA from single cells can be obtained as described in methods for preparing cDNA libraries from single cells, such as those described in
5 Dulac, C. (1998) Curr. Top. Dev. Biol. 36, 245 and Jena et al. (1996) J. Immunol. Methods 190:199. Care to avoid RNA degradation must be taken, e.g., by inclusion of RNAsin.

The RNA sample can then be enriched in particular species. In one embodiment, poly(A)+ RNA is isolated from the RNA sample. In general, such purification takes advantage of the poly-A tails on mRNA. In particular and as noted above, poly-T oligonucleotides may be
10 immobilized within on a solid support to serve as affinity ligands for mRNA. Kits for this purpose are commercially available, e.g., the MessageMaker kit (Life Technologies, Grand Island, NY).

In a preferred embodiment, the RNA population is enriched in sequences of interest, such as those of genes characteristic of R.A. Enrichment can be undertaken, e.g., by primer-specific cDNA synthesis, or multiple rounds of linear amplification based on cDNA synthesis and template-directed *in vitro* transcription (*see, e.g.*, Wang et al. (1989) PNAS 86, 9717; Dulac et al., *supra*, and Jena et al., *supra*).

The population of RNA, enriched or not in particular species or sequences, can further be amplified. Such amplification is particularly important when using RNA from a single or a few
20 cells. A variety of amplification methods are suitable for use in the methods of the invention, including, e.g., PCR; ligase chain reaction (LCR) (*see, e.g.*, Wu and Wallace, Genomics 4, 560 (1989), Landegren et al., Science 241, 1077 (1988)); self-sustained sequence replication (SSR) (*see, e.g.*, Guatelli et al., Proc. Nat. Acad. Sci. USA, 87, 1874 (1990)); nucleic acid based sequence amplification (NASBA) and transcription amplification (*see, e.g.*, Kwoh et al., Proc.
25 Natl. Acad. Sci. USA 86, 1173 (1989)). For PCR technology, *see, e.g.*, PCR Technology: Principles and Applications for DNA Amplification (ed. H. A. Erlich, Freeman Press, N.Y., N.Y., 1992); PCR Protocols: A Guide to Methods and applications (eds. Innis, et al., Academic Press, San Diego, Calif., 1990); Mattila et al., Nucleic Acids Res. 19, 4967 (1991); Eckert et al., PCR Methods and Applications 1, 17 (1991); PCR (eds. McPherson et al., IRL Press, Oxford);
30 and U.S. Pat. No. 4,683,202. Methods of amplification are described, e.g., in Ohyama et al. (2000) BioTechniques 29:530; Luo et al. (1999) Nat. Med. 5, 117; Hegde et al. (2000)

BioTechniques 29:548; Kacharina et al. (1999) Meth. Enzymol. 303:3; Livesey et al. (2000) Curr. Biol. 10:301; Spirin et al. (1999) Invest. Ophthalmol. Vis. Sci. 40:3108; and Sakai et al. (2000) Anal. Biochem. 287:32. RNA amplification and cDNA synthesis can also be conducted in cells *in situ* (see, e.g., Eberwine et al. (1992) PNAS 89:3010).

5 One of skill in the art will appreciate that whatever amplification method is used, if a quantitative result is desired, care must be taken to use a method that maintains or controls for the relative frequencies of the amplified nucleic acids to achieve quantitative amplification. Methods of "quantitative" amplification are well known to those of skill in the art. For example, quantitative PCR involves simultaneously co-amplifying a known quantity of a control sequence
10 using the same primers. This provides an internal standard that may be used to calibrate the PCR reaction. A high density array may then include probes specific to the internal standard for quantification of the amplified nucleic acid.

One preferred internal standard is a synthetic AW106 cRNA. The AW106 ERNA is combined with RNA isolated from the sample according to standard techniques known to those of skilled in the art. The RNA is then reverse transcribed using a reverse transcriptase to provide copy DNA. The cDNA sequences are then amplified (e.g., by PCR) using labeled primers. The amplification products are separated, typically by electrophoresis, and the amount of radioactivity (proportional to the amount of amplified product) is determined. The amount of mRNA in the sample is then calculated by comparison with the signal produced by the known
20 AW106 RNA standard. Detailed protocols for quantitative PCR are provided in PCR Protocols, A Guide to Methods and Applications, Innis et al., Academic Press, Inc. N.Y., (1990).

In a preferred embodiment, a sample mRNA is reverse transcribed with a reverse transcriptase and a primer consisting of oligo(dT) and a sequence encoding the phage T7 promoter to provide single stranded DNA template. The second DNA strand is polymerized
25 using a DNA polymerase. After synthesis of double-stranded cDNA, T7 RNA polymerase is added and RNA is transcribed from the cDNA template. Successive rounds of transcription from each single cDNA template results in amplified RNA. Methods of *in vitro* polymerization are well known to those of skill in the art (see, e.g., Sambrook, (*supra*) and this particular method is described in detail by Van Gelder, et al., Proc. Natl. Acad. Sci. USA, 87: 1663-1667 (1990) who
30 demonstrate that *in vitro* amplification according to this method preserves the relative frequencies of the various RNA transcripts). Moreover, Eberwine et al. Proc. Natl. Acad. Sci.

USA, 89: 3010-3014 provide a protocol that uses two rounds of amplification via *in vitro* transcription to achieve greater than 10^6 fold amplification of the original starting material, thereby permitting expression monitoring even where biological samples are limited.

It will be appreciated by one of skill in the art that the direct transcription method described above provides an antisense (aRNA) pool. Where antisense RNA is used as the target nucleic acid, the oligonucleotide probes provided in the array are chosen to be complementary to subsequences of the antisense nucleic acids. Conversely, where the target nucleic acid pool is a pool of sense nucleic acids, the oligonucleotide probes are selected to be complementary to subsequences of the sense nucleic acids. Finally, where the nucleic acid pool is double stranded, the probes may be of either sense as the target nucleic acids include both sense and antisense strands.

(ii) *Labeling of the nucleic acids to be analyzed*

Generally, the target molecules will be labeled to permit detection of hybridization of target molecules to a microarray. By "labeled" is meant that the probe comprises a member of a signal producing system and is thus detectable, either directly or through combined action with one or more additional members of a signal producing system. Examples of directly detectable labels include isotopic and fluorescent moieties incorporated into, usually covalently bonded to, a moiety of the probe, such as a nucleotide monomeric unit, e.g. dNMP of the primer, or a photoactive or chemically active derivative of a detectable label which can be bound to a functional moiety of the probe molecule.

Nucleic acids can be labeled after or during enrichment and/or amplification of RNAs. For example, labeled cDNA can be prepared from mRNA by oligo dT-primed or random-primed reverse transcription, both of which are well known in the art (*see, e.g.*, Klug and Berger, 1987, Methods Enzymol. 152:316-325). Reverse transcription may be carried out in the presence of a dNTP conjugated to a detectable label, most preferably a fluorescently labeled dNTP. Alternatively, isolated mRNA can be converted to labeled antisense RNA synthesized by *in vitro* transcription of double-stranded cDNA in the presence of labeled dNTPs (Lockhart et al., 1996, Expression monitoring by hybridization to high-density oligonucleotide arrays, Nature Biotech. 14:1675). In alternative embodiments, the cDNA or RNA probe can be synthesized in the absence of detectable label and may be labeled subsequently, e.g., by incorporating biotinylated dNTPs or rNTP, or some similar means (e.g., photo-cross-linking a psoralen derivative of biotin

to RNAs), followed by addition of labeled streptavidin (e.g., phycoerythrin-conjugated streptavidin) or the equivalent.

In one embodiment, labeled cDNA is synthesized by incubating a mixture containing RNA and 0.5 mM dGTP, dATP and dCTP plus 0.1 mM dTTP plus fluorescent deoxyribonucleotides (e.g., 0.1 mM Rhodamine 110 UTP (Perkin Elmer Cetus) or 0.1 mM Cy3 dUTP (Amersham)) with reverse transcriptase (e.g., SuperScript.TM.II, LTI Inc.) at 42 °C for 60 min.

Fluorescent moieties or labels of interest include coumarin and its derivatives, e.g. 7-amino-4-methylcoumarin, aminocoumarin, bodipy dyes, such as Bodipy FL, cascade blue, fluorescein and its derivatives, e.g. fluorescein isothiocyanate, Oregon green, rhodamine dyes, e.g. Texas red, tetramethylrhodamine, eosins and erythrosins, cyanine dyes, e.g. Cy2, Cy3, Cy3.5, Cy5, Cy5.5, Cy7, FluorX, macrocyclic chelates of lanthanide ions, e.g. quantum dyeTM, fluorescent energy transfer dyes, such as thiazole orange-ethidium heterodimer, TOTAB, dansyl, etc. Individual fluorescent compounds which have functionalities for linking to an element desirably detected in an apparatus or assay of the invention, or which can be modified to incorporate such functionalities include, e.g., dansyl chloride; fluoresceins such as 3,6-dihydroxy-9-phenylxanthidrol; rhodamineisothiocyanate; N-phenyl 1-amino-8-sulfonatonaphthalene; N-phenyl 2-amino-6-sulfonatonaphthalene; 4-acetamido-4-isothiocyanato-stilbene-2,2'-disulfonic acid; pyrene-3-sulfonic acid; 2-toluidinonaphthalene-6-sulfonate; N-phenyl-N-methyl-2-aminoaphthalene-6-sulfonate; ethidium bromide; stebrine; auromine-0,2-(9'-anthroyl)palmitate; dansyl phosphatidylethanolamine; N,N'-dioctadecyl oxacarbocyanine; N,N'-dihexyl oxacarbocyanine; merocyanine, 4-(3'-pyrenyl)stearate; d-3-aminodesoxy-equilenin; 12-(9'-anthroyl)stearate; 2-methylanthracene; 9-vinyanthracene; 2,2'(vinylene-p-phenylene)bisbenzoxazole; p-bis(2-methyl-5-phenyl-oxazolyl)benzene; 6-dimethylamino-1,2-benzophenazin; retinol; bis(3'-aminopyridinium) 1,10-decandiyl diiodide; sulfonaphthylhydrazine of hellibrienin; chlorotetracycline; N-(7-dimethylamino-4-methyl-2-oxo-3-chromenyl)maleimide; N-(p-(2benzimidazolyl)-phenyl)maleimide; N-(4-fluoranthyl)maleimide; bis(homovanillic acid); resazurin; 4-chloro-7-nitro-2,1,3-benzooxadiazole; merocyanine 540; resorufin; rose bengal; and 2,4-diphenyl-3(2H)-furanone. (see, e.g., Kricka, 1992, Nonisotopic DNA Probe Techniques, Academic Press San Diego, Calif.). Many fluorescent tags are commercially available from SIGMA chemical company

(Saint Louis, Mo.), Amersham, Molecular Probes, R&D systems (Minneapolis, Minn.), Pharmacia LKB Biotechnology (Piscataway, N.J.), CLONTECH Laboratories, Inc. (Palo Alto, Calif.), Chem Genes Corp., Aldrich Chemical Company (Milwaukee, Wis.), Glen Research, Inc., GIBCO BRL Life Technologies, Inc. (Gaithersburg, Md.), Fluka Chemica-Biochemika Analytika (Fluka Chemie AG, Buchs, Switzerland), and Applied Biosystems (Foster City, Calif.) as well as other commercial sources known to one of skill.

Chemiluminescent labels include luciferin and 2,3-dihydrophthalazinediones, e.g., luminol.

Isotopic moieties or labels of interest include ^{32}P , ^{33}P , ^{35}S , ^{125}I , ^2H , ^{14}C , and the like (*see* Zhao et al., 1995, High density cDNA filter analysis: a novel approach for large-scale, quantitative analysis of gene expression, *Gene* 156:207; Pietu et al., 1996, Novel gene transcripts preferentially expressed in human muscles revealed by quantitative hybridization of a high density cDNA array, *Genome Res.* 6:492).

Labels may also be members of a signal producing system that act in concert with one or more additional members of the same system to provide a detectable signal. Illustrative of such labels are members of a specific binding pair, such as ligands, e.g. biotin, fluorescein, digoxigenin, antigen, polyvalent cations, chelator groups and the like, where the members specifically bind to additional members of the signal producing system, where the additional members provide a detectable signal either directly or indirectly, e.g. antibody conjugated to a fluorescent moiety or an enzymatic moiety capable of converting a substrate to a chromogenic product, e.g. alkaline phosphatase conjugate antibody and the like.

Additional labels of interest include those that provide for signal only when the probe with which they are associated is specifically bound to a target molecule, where such labels include: "molecular beacons" as described in Tyagi & Kramer, *Nature Biotechnology* (1996) 14:303 and EP 0 070 685 B1. Other labels of interest include those described in U.S. Pat. No. 5,563,037; WO 97/17471 and WO 97/17076.

In some cases, hybridized target nucleic acids may be labeled following hybridization. For example, where biotin labeled dNTPs are used in, e.g., amplification or transcription, streptavidin linked reporter groups may be used to label hybridized complexes.

In other embodiments, the target nucleic acid is not labeled. In this case, hybridization can be determined, e.g., by plasmon resonance, as described, e.g., in Thiel et al. (1997) Anal. Chem. 69:4948.

In one embodiment, a plurality (e.g., 2, 3, 4, 5 or more) of sets of target nucleic acids are labeled and used in one hybridization reaction ("multiplex" analysis). For example, one set of nucleic acids may correspond to RNA from one cell or tissue sample and another set of nucleic acids may correspond to RNA from another cell or tissue sample. The plurality of sets of nucleic acids can be labeled with different labels, e.g., different fluorescent labels which have distinct emission spectra so that they can be distinguished. The sets can then be mixed and hybridized simultaneously to one microarray.

For example, the two different cells can be a diseased cell of a patient having R.A. and a counterpart normal cell. Alternatively, the two different cells can be a diseased cell of a patient having R.A. and a diseased cell of a patient suspected of having R.A. In another embodiment, one biological sample is exposed to a drug and another biological sample of the same type is not exposed to the drug. The cDNA derived from each of the two cell types are differently labeled so that they can be distinguished. In one embodiment, for example, cDNA from a diseased cell is synthesized using a fluorescein-labeled dNTP, and cDNA from a second cell, i.e., the normal cell, is synthesized using a rhodamine-labeled dNTP. When the two cDNAs are mixed and hybridized to the microarray, the relative intensity of signal from each cDNA set is determined for each site on the array, and any relative difference in abundance of a particular mRNA detected.

In the example described above, the cDNA from the diseased cell will fluoresce green when the fluorophore is stimulated and the cDNA from the cell of a subject suspected of having R.A. will fluoresce red. As a result, if the two cells are essentially the same, the particular mRNA will be equally prevalent in both cells and, upon reverse transcription, red-labeled and green-labeled cDNA will be equally prevalent. When hybridized to the microarray, the binding site(s) for that species of RNA will emit wavelengths characteristic of both fluorophores (and appear brown in combination). In contrast, if the two cells are different, the ratio of green to red fluorescence will be different.

The use of a two-color fluorescence labeling and detection scheme to define alterations in gene expression has been described, e.g., in Shena et al., 1995, Quantitative monitoring of gene

expression patterns with a complementary DNA microarray, Science 270:467-470. An advantage of using cDNA labeled with two different fluorophores is that a direct and internally controlled comparison of the mRNA levels corresponding to each arrayed gene in two cell states can be made, and variations due to minor differences in experimental conditions (e.g., hybridization conditions) will not affect subsequent analyses.

Examples of distinguishable labels for use when hybridizing a plurality of target nucleic acids to one array are well known in the art and include: two or more different emission wavelength fluorescent dyes, like Cy3 and Cy5, combination of fluorescent proteins and dyes, like phicoerythrin and Cy5, two or more isotopes with different energy of emission, like ^{32}P and ^{33}P , gold or silver particles with different scattering spectra, labels which generate signals under different treatment conditions, like temperature, pH, treatment by additional chemical agents, etc., or generate signals at different time points after treatment. Using one or more enzymes for signal generation allows for the use of an even greater variety of distinguishable labels, based on different substrate specificity of enzymes (alkaline phosphatase/peroxidase).

Further, it is preferable in order to reduce experimental error to reverse the fluorescent labels in two-color differential hybridization experiments to reduce biases peculiar to individual genes or array spot locations. In other words, it is preferable to first measure gene expression with one labeling (e.g., labeling nucleic acid from a first cell with a first fluorochrome and nucleic acid from a second cell with a second fluorochrome) of the mRNA from the two cells being measured, and then to measure gene expression from the two cells with reversed labeling (e.g., labeling nucleic acid from the first cell with the second fluorochrome and nucleic acid from the second cell with the first fluorochrome). Multiple measurements over exposure levels and perturbation control parameter levels provide additional experimental error control.

The quality of labeled nucleic acids can be evaluated prior to hybridization to an array. For example, a sample of the labeled nucleic acids can be hybridized to probes derived from the 5', middle and 3' portions of genes known to be or suspected to be present in the nucleic acid sample. This will be indicative as to whether the labeled nucleic acids are full length nucleic acids or whether they are degraded. In one embodiment, the GeneChip[®] Test3 Array from Affymetrix (Santa Clara, CA) can be used for that purpose. This array contains probes representing a subset of characterized genes from several organisms including mammals. Thus,

the quality of a labeled nucleic acid sample can be determined by hybridization of a fraction of the sample to an array, such as the GeneChip[®] Test3 Array from Affymetrix (Santa Clara, CA).

(iii) *Exemplary arrays*

Preferred arrays, e.g., microarrays, for use according to the invention include one or more probes of genes which are up- or down-regulated in R.A., such as one or more genes listed in any of Tables 1-5 or one or more genes characteristic of R.A. In a preferred embodiment, the array comprises probes corresponding to one or more of genes selected from the group consisting of genes which are up-regulated in R.A., e.g., genes selected from the group consisting of SOCS3 (CISH3); RAGE (AGER); LST-1 (LY117); SAA 1-3; HMG-1; S100 A8, A9, and A12; SLPI; GILZ; PTPN-18; GADD-45A and B; Legumain (PRSC1); FST1; Lcn2; GPI; SpiL; and TSG-6 and genes which are down-regulated, e.g., CAMK2B, PLA2G2A, GBAS and SOX15. The array may comprise probes corresponding to at least 10, preferably at least 20, at least 50, at least 100 or at least 1000 genes. The array may comprise probes corresponding to about 10%, 20%, 50%, 70%, 90% or 95% of the genes listed in any of Tables 1-5. The array may comprise probes corresponding to about 10%, 20%, 50%, 70%, 90% or 95% of the genes listed in any of Tables 1-5 whose expression is at least 2 fold, preferably at least 3 fold, more preferably at least 4 fold, 5 fold, 7 fold and most preferably at least about 10 fold higher in cells characteristic of R.A. relative to normal counterpart cells. One array that can be used is the array used and described in the Examples.

There can be one or more than one probe corresponding to each gene on a microarray. For example, a microarray may contain from 2 to 20 probes corresponding to one gene and preferably about 5 to 10. The probes may correspond to the full length RNA sequence or complement thereof of genes characteristic of R.A., or they may correspond to a portion thereof, which portion is of sufficient length for permitting specific hybridization. Such probes may comprise from about 50 nucleotides to about 100, 200, 500, or 1000 nucleotides or more than 1000 nucleotides. As further described herein, microarrays may contain oligonucleotide probes, consisting of about 10 to 50 nucleotides, preferably about 15 to 30 nucleotides and even more preferably 20-25 nucleotides. The probes are preferably single stranded. The probe will have sufficient complementarity to its target to provide for the desired level of sequence specific hybridization (*see below*).

Typically, the arrays used in the present invention will have a site density of greater than 100 different probes per cm^2 . Preferably, the arrays will have a site density of greater than $500/\text{cm}^2$, more preferably greater than about $1000/\text{cm}^2$, and most preferably, greater than about $10,000/\text{cm}^2$. Preferably, the arrays will have more than 100 different probes on a single substrate, more preferably greater than about 1000 different probes still more preferably, greater than about 10,000 different probes and most preferably, greater than 100,000 different probes on a single substrate.

Microarrays can be prepared by methods known in the art, as described below, or they can be custom made by companies, e.g., Affymetrix (Santa Clara, CA).

Generally, two types of microarrays can be used. These two types are referred to as "synthesis" and "delivery." In the synthesis type, a microarray is prepared in a step-wise fashion by the in situ synthesis of nucleic acids from nucleotides. With each round of synthesis, nucleotides are added to growing chains until the desired length is achieved. In the delivery type of microarray, preprepared nucleic acids are deposited onto known locations using a variety of delivery technologies. Numerous articles describe the different microarray technologies, e.g., Shena et al. (1998) Tibtech 16: 301; Duggan et al. (1999) Nat. Genet. 21:10; Bowtell et al. (1999) Nat. Genet. 21: 25.

One novel synthesis technology is that developed by Affymetrix (Santa Clara, CA), which combines photolithography technology with DNA synthetic chemistry to enable high density oligonucleotide microarray manufacture. Such chips contain up to 400,000 groups of oligonucleotides in an area of about 1.6 cm^2 . Oligonucleotides are anchored at the 3' end thereby maximizing the availability of single-stranded nucleic acid for hybridization. Generally such chips, referred to as "GeneChips[®]" contain several oligonucleotides of a particular gene, e.g., between 15-20, such as 16 oligonucleotides. Since Affymetrix (Santa Clara, CA) sells custom made microarrays, microarrays containing genes which are up- or down-regulated in R.A. can be ordered for purchase from Affymetrix (Santa Clara, CA).

Microarrays can also be prepared by mechanical microspotting, e.g., those commercialized at Synteni (Fremont, CA). According to these methods, small quantities of nucleic acids are printed onto solid surfaces. Microspotted arrays prepared at Synteni contain as many as 10,000 groups of cDNA in an area of about 3.6 cm^2 .

A third group of microarray technologies consist in the “drop-on-demand” delivery approaches, the most advanced of which are the ink-jetting technologies, which utilize piezoelectric and other forms of propulsion to transfer nucleic acids from miniature nozzles to solid surfaces. Inkjet technologies is developed at several centers including Incyte
5 Pharmaceuticals (Palo Alto, CA) and Protogene (Palo Alto, CA). This technology results in a density of 10,000 spots per cm². See also, Hughes et al. (2001) Nat. Biotechn. 19:342.

Arrays preferably include control and reference nucleic acids. Control nucleic acids are nucleic acids which serve to indicate that the hybridization was effective. For example, all Affymetrix (Santa Clara, CA) expression arrays contain sets of probes for several prokaryotic
10 genes, e.g., bioB, bioC and bioD from biotin synthesis of *E. coli* and cre from P1 bacteriophage. Hybridization to these arrays is conducted in the presence of a mixture of these genes or portions thereof, such as the mix provided by Affymetrix (Santa Clara, CA) to that effect (Part Number 900299), to thereby confirm that the hybridization was effective. Control nucleic acids included with the target nucleic acids can also be mRNA synthesized from cDNA clones by *in vitro*
15 transcription. Other control genes that may be included in arrays are polyA controls, such as dap, lys, phe, thr, and trp (which are included on Affymetrix GeneChips®)

Reference nucleic acids allow the normalization of results from one experiment to another, and to compare multiple experiments on a quantitative level. Exemplary reference
20 nucleic acids include housekeeping genes of known expression levels, e.g., GAPDH, hexokinase and actin.

Mismatch controls may also be provided for the probes to the target genes, for expression level controls or for normalization controls. Mismatch controls are oligonucleotide probes or other nucleic acid probes identical to their corresponding test or control probes except for the presence of one or more mismatched bases.

25 Arrays may also contain probes that hybridize to more than one allele of a gene. For example the array can contain one probe that recognizes allele 1 and another probe that recognizes allele 2 of a particular gene.

Microarrays can be prepared as follows. In one embodiment, an array of oligonucleotides is synthesized on a solid support. Exemplary solid supports include glass,
30 plastics, polymers, metals, metalloids, ceramics, organics, etc. Using chip masking technologies and photoprotective chemistry it is possible to generate ordered arrays of nucleic acid probes.

These arrays, which are known, e.g., as "DNA chips," or as very large scale immobilized polymer arrays ("VLSIPSTM" arrays) can include millions of defined probe regions on a substrate having an area of about 1 cm² to several cm², thereby incorporating sets of from a few to millions of probes (*see, e.g.*, U.S. Patent No. 5,631,734).

5 The construction of solid phase nucleic acid arrays to detect target nucleic acids is well described in the literature. *See*, Fodor et al. (1991) Science, 251: 767-777; Sheldon et al. (1993) Clinical Chemistry 39(4): 718-719; Kozal et al. (1996) Nature Medicine 2(7): 753-759 and Hubbell U.S. Pat. No. 5,571,639; Pinkel et al. PCT/US95/16155 (WO 96/17958); U.S. Pat. Nos. 5,677,195; 5,624,711; 5,599,695; 5,451,683; 5,424,186; 5,412,087; 5,384,261; 5,252,743 and 10 5,143,854; PCT Patent Publication Nos. 92/10092 and 93/09668; and PCT WO 97/10365. In brief, a combinatorial strategy allows for the synthesis of arrays containing a large number of probes using a minimal number of synthetic steps. For instance, it is possible to synthesize and attach all possible DNA 8 mer oligonucleotides (48, or 65,536 possible combinations) using only 32 chemical synthetic steps. In general, VLSIPSTM procedures provide a method of producing 4n different oligonucleotide probes on an array using only 4n synthetic steps (*see, e.g.*, U.S. Pat. No. 5,631,734 5; 143,854 and PCT Patent Publication Nos. WO 90/15070; WO 95/11995 and WO 92/10092).

15 Light-directed combinatorial synthesis of oligonucleotide arrays on a glass surface can be performed with automated phosphoramidite chemistry and chip masking techniques similar to photoresist technologies in the computer chip industry. Typically, a glass surface is derivatized with a silane reagent containing a functional group, e.g., a hydroxyl or amine group blocked by a photolabile protecting group. Photolysis through a photolithographic mask is used selectively to 20 expose functional groups which are then ready to react with incoming 5'-photoprotected nucleoside phosphoramidites. The phosphoramidites react only with those sites which are illuminated (and thus exposed by removal of the photolabile blocking group). Thus, the phosphoramidites only add to those areas selectively exposed from the preceding step. These steps are repeated until the desired array of sequences have been synthesized on the solid surface.

25 Algorithms for design of masks to reduce the number of synthesis cycles are described by Hubbel et al., U.S. Pat. No. 5,571,639 and U.S. Pat. No. 5,593,839. A computer system may be used to select nucleic acid probes on the substrate and design the layout of the array as described in U.S. Pat. No. 5,571,639.

Another method for synthesizing high density arrays is described in U.S. Patent No. 6,083,697. This method utilizes a novel chemical amplification process using a catalyst system which is initiated by radiation to assist in the synthesis the polymer sequences. Such methods include the use of photosensitive compounds which act as catalysts to chemically alter the synthesis intermediates in a manner to promote formation of polymer sequences. Such photosensitive compounds include what are generally referred to as radiation-activated catalysts (RACs), and more specifically photo activated catalysts (PACs). The RACs can by themselves chemically alter the synthesis intermediate or they can activate an autocatalytic compound which chemically alters the synthesis intermediate in a manner to allow the synthesis intermediate to chemically combine with a later added synthesis intermediate or other compound.

Arrays can also be synthesized in a combinatorial fashion by delivering monomers to cells of a support by mechanically constrained flowpaths. *See* Winkler et al., EP 624,059. Arrays can also be synthesized by spotting monomers reagents on to a support using an ink jet printer. *See id.* and Pease et al., EP 728,520.

cDNA probes can be prepared according to methods known in the art and further described herein, e.g., reverse-transcription PCR (RT-PCR) of RNA using sequence specific primers. Oligonucleotide probes can be synthesized chemically. Sequences of the genes or cDNA from which probes are made can be obtained, e.g., from GenBank, other public databases or publications.

Nucleic acid probes can be natural nucleic acids, chemically modified nucleic acids, e.g., composed of nucleotide analogs, as long as they have activated hydroxyl groups compatible with the linking chemistry. The protective groups can, themselves, be photolabile. Alternatively, the protective groups can be labile under certain chemical conditions, e.g., acid. In this example, the surface of the solid support can contain a composition that generates acids upon exposure to light. Thus, exposure of a region of the substrate to light generates acids in that region that remove the protective groups in the exposed region. Also, the synthesis method can use 3'-protected 5'-O-phosphoramidite-activated deoxynucleoside. In this case, the oligonucleotide is synthesized in the 5' to 3' direction, which results in a free 5' end.

Oligonucleotides of an array can be synthesized using a 96 well automated multiplex oligonucleotide synthesizer (A.M.O.S.) that is capable of making thousands of oligonucleotides (Lashkari et al. (1995) PNAS 93: 7912) can be used.

It will be appreciated that oligonucleotide design is influenced by the intended application. For example, it may be desirable to have similar melting temperatures for all of the probes. Accordingly, the length of the probes are adjusted so that the melting temperatures for all of the probes on the array are closely similar (it will be appreciated that different lengths for different probes may be needed to achieve a particular T_m where different probes have different GC contents). Although melting temperature is a primary consideration in probe design, other factors are optionally used to further adjust probe construction, such as selecting against primer self-complementarity and the like.

Arrays, e.g., microarrays, may conveniently be stored following fabrication or purchase for use at a later time. Under appropriate conditions, the subject arrays are capable of being stored for at least about 6 months and may be stored for up to one year or longer. Arrays are generally stored at temperatures between about -20°C to room temperature, where the arrays are preferably sealed in a plastic container, e.g. bag, and shielded from light.

(iv) *Hybridization of the target nucleic acids to the microarray*

The next step is to contact the target nucleic acids with the array under conditions sufficient for binding between the target nucleic acids and the probes of the array. In a preferred embodiment, the target nucleic acids will be contacted with the array under conditions sufficient for hybridization to occur between the target nucleic acids and probes on the microarray, where the hybridization conditions will be selected in order to provide for the desired level of hybridization specificity.

Contact of the array and target nucleic acids involves contacting the array with an aqueous medium comprising the target nucleic acids. Contact may be achieved in a variety of different ways depending on specific configuration of the array. For example, where the array simply comprises the pattern of size separated probes on the surface of a "plate-like" rigid substrate, contact may be accomplished by simply placing the array in a container comprising the target nucleic acid solution, such as a polyethylene bag, and the like. In other embodiments where the array is entrapped in a separation media bounded by two rigid plates, the opportunity exists to deliver the target nucleic acids via electrophoretic means. Alternatively, where the array is incorporated into a biochip device having fluid entry and exit ports, the target nucleic acid solution can be introduced into the chamber in which the pattern of target molecules is presented through the entry port, where fluid introduction could be performed manually or with

an automated device. In multiwell embodiments, the target nucleic acid solution will be introduced in the reaction chamber comprising the array, either manually, e.g. with a pipette, or with an automated fluid handling device.

Contact of the target nucleic acid solution and the probes will be maintained for a sufficient period of time for binding between the target and the probe to occur. Although dependent on the nature of the probe and target, contact will generally be maintained for a period of time ranging from about 10 min to 24 hrs, usually from about 30 min to 12 hrs and more usually from about 1 hr to 6 hrs.

When using commercially available microarrays, adequate hybridization conditions are provided by the manufacturer. When using non-commercial microarrays, adequate hybridization conditions can be determined based on the following hybridization guidelines, as well as on the hybridization conditions described in the numerous published articles on the use of microarrays.

Nucleic acid hybridization and wash conditions are optimally chosen so that the probe “specifically binds” or “specifically hybridizes” to a specific array site, i.e., the probe hybridizes, duplexes or binds to a sequence array site with a complementary nucleic acid sequence but does not hybridize to a site with a non-complementary nucleic acid sequence. As used herein, one polynucleotide sequence is considered complementary to another when, if the shorter of the polynucleotides is less than or equal to 25 bases, there are no mismatches using standard base-pairing rules or, if the shorter of the polynucleotides is longer than 25 bases, there is no more than a 5% mismatch. Preferably, the polynucleotides are perfectly complementary (no mismatches). It can easily be demonstrated that specific hybridization conditions result in specific hybridization by carrying out a hybridization assay including negative controls.

Hybridization is carried out in conditions permitting essentially specific hybridization. The length of the probe and GC content will determine the T_m of the hybrid, and thus the hybridization conditions necessary for obtaining specific hybridization of the probe to the template nucleic acid. These factors are well known to a person of skill in the art, and can also be tested in assays. An extensive guide to the hybridization of nucleic acids is found in Tijssen (1993), “Laboratory Techniques in biochemistry and molecular biology-hybridization with nucleic acid probes.” Generally, stringent conditions are selected to be about 5°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH. The T_m is the temperature (under defined ionic strength and pH) at which 50% of the target sequence

hybridizes to a perfectly matched probe. Highly stringent conditions are selected to be equal to the T_m point for a particular probe. Sometimes the term “T_d” is used to define the temperature at which at least half of the probe dissociates from a perfectly matched target nucleic acid. In any case, a variety of estimation techniques for estimating the T_m or T_d are available, and generally described in Tijssen, *supra*. Typically, G-C base pairs in a duplex are estimated to contribute about 3°C to the T_m, while A-T base pairs are estimated to contribute about 2°C, up to a theoretical maximum of about 80-100°C. However, more sophisticated models of T_m and T_d are available and appropriate in which G-C stacking interactions, solvent effects, the desired assay temperature and the like are taken into account. For example, probes can be designed to have a dissociation temperature (T_d) of approximately 60°C, using the formula: $T_d = (((((3 \times \#GC) + (2 \times \#AT)) \times 37) - 562) / \#bp) - 5$; where #GC, #AT, and #bp are the number of guanine-cytosine base pairs, the number of adenine-thymine base pairs, and the number of total base pairs, respectively, involved in the annealing of the probe to the template DNA.

The stability difference between a perfectly matched duplex and a mismatched duplex, particularly if the mismatch is only a single base, can be quite small, corresponding to a difference in T_m between the two of as little as 0.5 degrees. See Tibanyenda, N. et al., *Eur. J. Biochem.* 139:19 (1984) and Ebel, S. et al., *Biochem.* 31:12083 (1992). More importantly, it is understood that as the length of the homology region increases, the effect of a single base mismatch on overall duplex stability decreases.

Theory and practice of nucleic acid hybridization is described, e.g., in S. Agrawal (ed.) *Methods in Molecular Biology*, volume 20; and Tijssen (1993) *Laboratory Techniques in biochemistry and molecular biology-hybridization with nucleic acid probes*, e.g., part I chapter 2 “Overview of principles of hybridization and the strategy of nucleic acid probe assays”, Elsevier, New York provide a basic guide to nucleic acid hybridization.

Certain microarrays are of “active” nature, i.e., they provide independent electronic control over all aspects of the hybridization reaction (or any other affinity reaction) occurring at each specific microlocation. These devices provide a new mechanism for affecting hybridization reactions which is called electronic stringency control (ESC). Such active devices can electronically produce “different stringency conditions” at each microlocation. Thus, all hybridizations can be carried out optimally in the same bulk solution. These arrays are described in U.S. Patent No. 6,051,380 by Sosnowski et al.

In a preferred embodiment, background signal is reduced by the use of a detergent (e.g., C-TAB) or a blocking reagent (e.g., sperm DNA, cot-1 DNA, etc.) during the hybridization to reduce non-specific binding. In a particularly preferred embodiment, the hybridization is performed in the presence of about 0.5 mg/ml DNA (e.g., herring sperm DNA). The use of blocking agents in hybridization is well known to those of skill in the art (see, e.g., Chapter 8 in Laboratory Techniques in Biochemistry and Molecular Biology, Vol. 24: Hybridization With Nucleic Acid Probes, P. Tijssen, ed. Elsevier, N.Y., (1993)).

The method may or may not further comprise a non-bound label removal step prior to the detection step, depending on the particular label employed on the target nucleic acid. For example, in certain assay formats (e.g., "homogenous assay formats") a detectable signal is only generated upon specific binding of target to probe. As such, in these assay formats, the hybridization pattern may be detected without a non-bound label removal step. In other embodiments, the label employed will generate a signal whether or not the target is specifically bound to its probe. In such embodiments, the non-bound labeled target is removed from the support surface. One means of removing the non-bound labeled target is to perform the well known technique of washing, where a variety of wash solutions and protocols for their use in removing non-bound label are known to those of skill in the art and may be used. Alternatively, non-bound labeled target can be removed by electrophoretic means.

Where all of the target sequences are detected using the same label, different arrays will be employed for each physiological source (where different could include using the same array at different times). The above methods can be varied to provide for multiplex analysis, by employing different and distinguishable labels for the different target populations (representing each of the different physiological sources being assayed). According to this multiplex method, the same array is used at the same time for each of the different target populations.

In another embodiment, hybridization is monitored in real time using a charge-coupled device (CCD) imaging camera (Guschin et al. (1997) Anal. Biochem. 250:203). Synthesis of arrays on optical fibre bundles allows easy and sensitive reading (Healy et al. (1997) Anal. Biochem. 251:270). In another embodiment, real time hybridization detection is carried out on microarrays without washing using evanescent wave effect that excites only fluorophores that are bound to the surface (see, e.g., Stimpson et al. (1995) PNAS 92:6379).

(v) *Detection of hybridization and analysis of results*

The above steps result in the production of hybridization patterns of target nucleic acid on the array surface. These patterns may be visualized or detected in a variety of ways, with the particular manner of detection being chosen based on the particular label of the target nucleic acid. Representative detection means include scintillation counting, autoradiography, fluorescence measurement, colorimetric measurement, light emission measurement, light scattering, and the like.

One method of detection includes an array scanner that is commercially available from Affymetrix (Santa Clara, CA), e.g., the 417TM Arrayer, the 418TM Array Scanner, or the Agilent GeneArrayTM Scanner. This scanner is controlled from the system computer with a Windows^R interface and easy-to-use software tools. The output is a 16-bit.tif file that can be directly imported into or directly read by a variety of software applications. Preferred scanning devices are described in, e.g., U.S. Pat. Nos. 5,143,854 and 5,424,186.

When fluorescently labeled probes are used, the fluorescence emissions at each site of a transcript array can be detected by scanning confocal laser microscopy. In one embodiment, a separate scan, using the appropriate excitation line, is carried out for each of the two fluorophores used. Alternatively, a laser can be used that allows simultaneous specimen illumination at wavelengths specific to the two fluorophores and emissions from the two fluorophores can be analyzed simultaneously (*see* Shalon et al., 1996, A DNA microarray system for analyzing complex DNA samples using two-color fluorescent probe hybridization, *Genome Research* 6:639-645). In a preferred embodiment, the arrays are scanned with a laser fluorescent scanner with a computer controlled X-Y stage and a microscope objective. Sequential excitation of the two fluorophores can be achieved with a multi-line, mixed gas laser and the emitted light is split by wavelength and detected with two photomultiplier tubes. In one embodiment in which fluorescent target nucleic acids are used, the arrays may be scanned using lasers to excite fluorescently labeled targets that have hybridized to regions of probe arrays, which can then be imaged using charged coupled devices ("CCDs") for a wide field scanning of the array. Fluorescence laser scanning devices are described, e.g., in Schena et al., 1996, *Genome Res.* 6:639-645. Alternatively, the fiber-optic bundle described by Ferguson et al., 1996, *Nature Biotech.* 14:1681-1684, may be used to monitor mRNA abundance levels.

Following the data gathering operation, the data will typically be reported to a data analysis operation. To facilitate the sample analysis operation, the data obtained by the reader

from the device will typically be analyzed using a digital computer. Typically, the computer will be appropriately programmed for receipt and storage of the data from the device, as well as for analysis and reporting of the data gathered, e.g., subtraction of the background, deconvolution multi-color images, flagging or removing artifacts, verifying that controls have performed properly, normalizing the signals, interpreting fluorescence data to determine the amount of hybridized target, normalization of background and single base mismatch hybridizations, and the like. In a preferred embodiment, a system comprises a search function that allows one to search for specific patterns, e.g., patterns relating to differential gene expression, e.g., between the expression profile of a cell of R.A. and the expression profile of a counterpart normal cell in a subject. A system preferably allows one to search for patterns of gene expression between more than two samples.

A desirable system for analyzing data is a general and flexible system for the visualization, manipulation, and analysis of gene expression data. Such a system preferably includes a graphical user interface for browsing and navigating through the expression data, allowing a user to selectively view and highlight the genes of interest. The system also preferably includes sort and search functions and is preferably available for general users with PC, Mac or Unix workstations. Also preferably included in the system are clustering algorithms that are qualitatively more efficient than existing ones. The accuracy of such algorithms is preferably hierarchically adjustable so that the level of detail of clustering can be systematically refined as desired.

Various algorithms are available for analyzing the gene expression profile data, e.g., the type of comparisons to perform. In certain embodiments, it is desirable to group genes that are co-regulated. This allows the comparison of large numbers of profiles. A preferred embodiment for identifying such groups of genes involves clustering algorithms (for reviews of clustering algorithms, *see, e.g.*, Fukunaga, 1990, Statistical Pattern Recognition, 2nd Ed., Academic Press, San Diego; Everitt, 1974, Cluster Analysis, London: Heinemann Educ. Books; Hartigan, 1975, Clustering Algorithms, New York: Wiley; Sneath and Sokal, 1973, Numerical Taxonomy, Freeman; Anderberg, 1973, Cluster Analysis for Applications, Academic Press: New York).

Clustering analysis is useful in helping to reduce complex patterns of thousands of time curves into a smaller set of representative clusters. Some systems allow the clustering and viewing of genes based on sequences. Other systems allow clustering based on other

characteristics of the genes, e.g., their level of expression (*see, e.g.*, U.S. Patent No. 6,203,987). Other systems permit clustering of time curves (*see, e.g.* U.S. Patent No. 6,263,287). Cluster analysis can be performed using the hclust routine (*see, e.g.*, “hclust” routine from the software package S-Plus, MathSoft, Inc., Cambridge, Mass.).

5 In some specific embodiments, genes are grouped according to the degree of co-variation of their transcription, presumably co-regulation, as described in U.S. Patent No. 6,203,987. Groups of genes that have co-varying transcripts are termed “genesets.” Cluster analysis or other statistical classification methods can be used to analyze the co-variation of transcription of genes in response to a variety of perturbations, e.g. caused by a disease or a drug. In one specific
10 embodiment, clustering algorithms are applied to expression profiles to construct a “similarity tree” or “clustering tree” which relates genes by the amount of co-regulation exhibited. Genesets are defined on the branches of a clustering tree by cutting across the clustering tree at different levels in the branching hierarchy.

In some embodiments, a gene expression profile is converted to a projected gene expression profile. The projected gene expression profile is a collection of geneset expression values. The conversion is achieved, in some embodiments, by averaging the level of expression of the genes within each geneset. In some other embodiments, other linear projection processes may be used. The projection operation expresses the profile on a smaller and biologically more meaningful set of coordinates, reducing the effects of measurement errors by averaging them
15 over each cellular constituent sets and aiding biological interpretation of the profile.

Values that can be compared include gross expression levels; averages of expression levels, e.g., from different experiments, different samples from the same subject or samples from different subjects; and ratios of expression levels, e.g., between R.A. subjects and normal controls, between different R.A. subjects and isolated cell populations.

2.2. Other methods for determining gene expression levels

25 In certain embodiments, it is sufficient to determine the expression of one or only a few genes, as opposed to hundreds or thousands of genes. Although microarrays can be used in these embodiments, various other methods of detection of gene expression are available. This section
30 describes a few exemplary methods for detecting and quantifying mRNA or polypeptide encoded thereby. Where the first step of the methods includes isolation of mRNA from cells, this step



can be conducted as described above. Labeling of one or more nucleic acids can be performed as described above.

1 In one embodiment, mRNA obtained from a sample is reverse transcribed into a first
cDNA strand and subjected to PCR, e.g., RT-PCR. House keeping genes, or other genes whose
5 expression does not vary can be used as internal controls and controls across experiments.
Following the PCR reaction, the amplified products can be separated by electrophoresis and
detected. By using quantitative PCR, the level of amplified product will correlate with the level
of RNA that was present in the sample. The amplified samples can also be separated on a
agarose or polyacrylamide gel, transferred onto a filter, and the filter hybridized with a probe
10 specific for the gene of interest. Numerous samples can be analyzed simultaneously by
conducting parallel PCR amplification, e.g., by multiplex PCR.

1002345
1.3.1.7.0.9
20 A quantitative PCR technique that can be used is based on the use of TaqManTM probes.
Specific sequence detection occurs by amplification of target sequences in the PE Applied
Biosystems 7700 Sequence Detection System in the presence of an oligonucleotide probe labeled
at the 5' and 3' ends with a reporter and quencher fluorescent dye, respectively (FQ probe), which
anneals between the two PCR primers. Only specific product will be detected when the probe is
bound between the primers. As PCR amplification proceeds, the 5'-nuclease activity of Taq
polymerase initially cleaves the reporter dye from the probe. The signal generated when the
reporter dye is physically separated from the quencher dye is detected by measuring the signal
with an attached CCD camera. Each signal generated equals one probe cleaved which
corresponds to amplification of one target strand. PCR reactions may be set up using the PE
Applied Biosystem TaqMan PCR Core Reagent Kit according to the instructions supplied. This
technique is further described, e.g., in U.S. Patent 6,326,462.

25 In another embodiment, mRNA levels is determined by dotblot analysis and related
methods (*see, e.g.,* G. A. Beltz et al., in *Methods in Enzymology*, Vol. 100, Part B, R. Wu, L.
Grossman, K. Moldave, Eds., Academic Press, New York, Chapter 19, pp. 266-308, 1985). In
one embodiment, a specified amount of RNA extracted from cells is blotted (i.e., non-covalently
bound) onto a filter, and the filter is hybridized with a probe of the gene of interest. Numerous
RNA samples can be analyzed simultaneously, since a blot can comprise multiple spots of RNA.
30 Hybridization is detected using a method that depends on the type of label of the probe. In
another dotblot method, one or more probes of one or more genes which are up- or down-

regulated in R.A. are attached to a membrane, and the membrane is incubated with labeled nucleic acids obtained from and optionally derived from RNA of a cell or tissue of a subject. Such a dotblot is essentially an array comprising fewer probes than a microarray.

“Dot blot” hybridization gained wide-spread use, and many versions were developed
5 (see, e.g., M. L. M. Anderson and B. D. Young, in *Nucleic Acid Hybridization-A Practical Approach*, B. D. Hames and S. J. Higgins, Eds., IRL Press, Washington D.C., Chapter 4, pp. 73-111, 1985).

Another format, the so-called “sandwich” hybridization, involves covalently attaching oligonucleotide probes to a solid support and using them to capture and detect multiple nucleic
10 acid targets (see, e.g., M. Ranki et al., *Gene*, 21, pp. 77-85, 1983; A. M. Palva, T. M. Ranki, and H. E. Soderlund, in UK Patent Application GB 2156074A, Oct. 2, 1985; T. M. Ranki and H. E. Soderlund in U.S. Pat. No. 4,563,419, Jan. 7, 1986; A. D. B. Malcolm and J. A. Langdale, in PCT WO 86/03782, Jul. 3, 1986; Y. Stabinsky, in U.S. Pat. No. 4,751,177, Jan. 14, 1988; T. H. Adams et al., in PCT WO 90/01564, Feb. 22, 1990; R. B. Wallace et al. 6 *Nucleic Acid Res.* 11, p. 3543, 1979; and B. J. Connor et al., 80 *Proc. Natl. Acad. Sci. USA* pp. 278-282, 1983). Multiplex versions of these formats are called “reverse dot blots.”

mRNA levels can also be determined by Northern blots. Specific amounts of RNA are separated by gel electrophoresis and transferred onto a filter which is then hybridized with a
15 probe corresponding to the gene of interest. This method, although more burdensome when numerous samples and genes are to be analyzed provides the advantage of being very accurate.
20

A preferred method for high throughput analysis of gene expression is the serial analysis of gene expression (SAGE) technique, first described in Velculescu et al. (1995) *Science* 270, 484-487. Among the advantages of SAGE is that it has the potential to provide detection of all genes expressed in a given cell type, provides quantitative information about the relative
25 expression of such genes, permits ready comparison of gene expression of genes in two cells, and yields sequence information that can be used to identify the detected genes. Thus far, SAGE methodology has proved itself to reliably detect expression of regulated and nonregulated genes in a variety of cell types (Velculescu et al. (1997) *Cell* 88, 243-251; Zhang et al. (1997) *Science* 276, 1268-1272 and Velculescu et al. (1999) *Nat. Genet.* 23, 387-388).

Techniques for producing and probing nucleic acids are further described, for example, in Sambrook *et al.*, "Molecular Cloning: A Laboratory Manual" (New York, Cold Spring Harbor Laboratory, 1989).

5 Alternatively, the level of expression of one or more genes which are up- or down-regulated in R.A. is determined by *in situ* hybridization. In one embodiment, a tissue sample is obtained from a subject, the tissue sample is sliced, and *in situ* hybridization is performed according to methods known in the art, to determine the level of expression of the genes of interest.

10 In other methods, the level of expression of a gene is detected by measuring the level of protein encoded by the gene. This can be done, e.g., by immunoprecipitation, ELISA, or immunohistochemistry using an agent, e.g., an antibody, that specifically detects the protein encoded by the gene. Other techniques include Western blot analysis. Immunoassays are commonly used to quantitate the levels of proteins in cell samples, and many other immunoassay techniques are known in the art. The invention is not limited to a particular assay procedure, and therefore is intended to include both homogeneous and heterogeneous procedures. Exemplary immunoassays which can be conducted according to the invention include fluorescence polarization immunoassay (FPIA), fluorescence immunoassay (FIA), enzyme immunoassay (EIA), nephelometric inhibition immunoassay (NIA), enzyme linked immunosorbent assay (ELISA), and radioimmunoassay (RIA). An indicator moiety, or label group, can be attached to
15 the subject antibodies and is selected so as to meet the needs of various uses of the method which are often dictated by the availability of assay equipment and compatible immunoassay procedures. General techniques to be used in performing the various immunoassays noted above are known to those of ordinary skill in the art.

25 In the case of polypeptides which are secreted from cells, the level of expression of these polypeptides can be measured in biological fluids.

2.3. Data analysis methods

30 Comparison of the expression levels of one or more genes which are up- or down-regulated in R.A. with reference expression levels, e.g., expression levels in cells characteristic of R.A. or in normal counterpart cells, is preferably conducted using computer systems. In one embodiment, one or more expression levels are obtained in two cells and these two sets of

expression levels are introduced into a computer system for comparison. In a preferred embodiment, one set of one or more expression levels is entered into a computer system for comparison with values that are already present in the computer system, or in computer-readable form that is then entered into the computer system.

5 In one embodiment, the invention provides a computer readable form of the gene expression profile data of the invention, or of values corresponding to the level of expression of at least one gene which is up- or down-regulated in R.A. The values can be mRNA expression levels obtained from experiments, e.g., microarray analysis. The values can also be mRNA levels normalized relative to a reference gene whose expression is constant in numerous cells
10 under numerous conditions, e.g., GAPDH. In other embodiments, the values in the computer are ratios of, or differences between, normalized or non-normalized mRNA levels in different samples.

The computer readable medium may comprise values of at least 2, at least 3, at least 5, 10, 20, 50, 100, 200, 500 or more genes, e.g., genes listed in Tables 1-5. In a preferred embodiment, the computer readable medium comprises at least one expression profile.

Gene expression data can be in the form of a table, such as an Excel table. The data can be alone, or it can be part of a larger database, e.g., comprising other expression profiles, e.g., publicly available database. The computer readable form can be in a computer. In another embodiment, the invention provides a computer displaying the gene expression profile data.

20 Although the invention provides methods in which the level of expression of a single gene can be compared in two or more cells or tissue samples, in a preferred embodiment, the level of expression of a plurality of genes is compared. For example, the level of expression of at least 2, at least 3, at least 5, 10, 20, 50, 100, 200, 500 or more genes, e.g., genes listed in Tables 1-5 can be compared. In a preferred embodiment, expression profiles are compared.

25 In one embodiment, the invention provides a method for determining the similarity between the level of expression of one or more genes which are up- or down-regulated in R.A. in a first cell, e.g., a cell of a subject, and that in a second cell. The method preferably comprises obtaining the level of expression of one or more genes which are up- or down-regulated in R.A. in a first cell and entering these values into a computer comprising (i) a database including
30 records comprising values corresponding to levels of expression of one or more genes which are up- or down-regulated in R.A. in a second cell, and (ii) processor instructions, e.g., a user

interface, capable of receiving a selection of one or more values for comparison purposes with data that is stored in the computer. The computer may further comprise a means for converting the comparison data into a diagram or chart or other type of output.

5 In another embodiment, values representing expression levels of one or more genes which are up- or down-regulated in R.A. are entered into a computer system which comprises one or more databases with reference expression levels obtained from more than one cell. For example, the computer may comprise expression data of diseased and normal cells. Instructions are provided to the computer, and the computer is capable of comparing the data entered with the data in the computer to determine whether the data entered is more similar to that of a normal
10 cell or to that of a diseased cell.

In another embodiment, the computer comprises values of expression levels in cells of subjects at different stages of R.A., and the computer is capable of comparing expression data entered into the computer with the data stored, and produce results indicating to which of the expression data in the computer, the one entered is most similar, such as to determine the stage of R.A. in the subject.

In yet another embodiment, the reference expression data in the computer are expression data from cells of R.A. of one or more subjects, which cells are treated *in vivo* or *in vitro* with a drug used for therapy of R.A. Upon entering of expression data of a cell of a subject treated *in vitro* or *in vivo* with the drug, the computer is instructed to compare the data entered with the data in the computer, and to provide results indicating whether the expression data input into the
20 computer are more similar to those of a cell of a subject that is responsive to the drug or more similar to those of a cell of a subject that is not responsive to the drug. Thus, the results indicate whether the subject is likely to respond to the treatment with the drug or unlikely to respond to it.

The reference expression data may also be from cells from subjects responding or not
25 responding to several different treatments, and the computer system indicates a preferred treatment for the subject. Accordingly, the invention provides a method for selecting a therapy for a patient having R.A., the method comprising: (i) providing the level of expression of one or more genes which are up- or down-regulated in R.A. in a diseased cell of the patient; (ii) providing a plurality of reference expression levels, each associated with a therapy, wherein the
30 subject expression levels and each reference expression level has a plurality of values, each value representing the level of expression of a gene that is up- or down-regulated in R.A.; and (iii)

selecting the reference expression levels most similar to the subject expression levels, to thereby select a therapy for said patient. In a preferred embodiment step (iii) is performed by a computer. The most similar reference profile may be selected by weighing a comparison value of the plurality using a weight value associated with the corresponding expression data.

5 In one embodiment, the invention provides a system that comprises a means for receiving gene expression data for one or a plurality of genes; a means for comparing the gene expression data from each of said one or plurality of genes to a common reference frame; and a means for presenting the results of the comparison. This system may further comprise a means for clustering the data.

10 In another embodiment, the invention provides a computer program for analyzing gene expression data comprising (i) a computer code that receives as input gene expression data for a plurality of genes and (ii) a computer code that compares said gene expression data from each of said plurality of genes to a common reference frame.

15 The invention also provides a machine-readable or computer-readable medium including program instructions for performing the following steps: (i) comparing a plurality of values corresponding to expression levels of one or more genes which are up- or down-regulated in R.A. in a query cell with a database including records comprising reference expression of one or more reference cells and an annotation of the type of cell; and (ii) indicating to which cell the query cell is most similar based on similarities of expression levels.

20 The relative levels of expression, e.g., abundance of an mRNA, in two biological samples can be scored as a perturbation (relative abundance difference) or as not perturbed (i.e., the relative abundance is the same). For example, a perturbation can be a difference in expression levels between the two sources of RNA of at least a factor of about 25% (RNA from one source is 25% more abundant in one source than the other source), more usually about 50%, even more often by a factor of about 2 (twice as abundant), 3 (three times as abundant) or 5 (five times as abundant). Perturbations can be used by a computer for calculating and expressing comparisons.

25 Preferably, in addition to identifying a perturbation as positive or negative, it is advantageous to determine the magnitude of the perturbation. This can be carried out, as noted above, by calculating the ratio of the emission of the two fluorophores used for differential labeling, or by analogous methods that will be readily apparent to those of skill in the art.

30

The computer readable medium may further comprise a pointer to a descriptor of the level of expression or expression profile, e.g., from which source it was obtained, e.g., from which patient it was obtained. A descriptor can reflect the stage of R.A., the therapy that the patient is undergoing or any other descriptions of the source of expression levels.

5 In operation, the means for receiving gene expression data, the means for comparing the gene expression data, the means for presenting, the means for normalizing, and the means for clustering within the context of the systems of the present invention can involve a programmed computer with the respective functionalities described herein, implemented in hardware or hardware and software; a logic circuit or other component of a programmed computer that
10 performs the operations specifically identified herein, dictated by a computer program; or a computer memory encoded with executable instructions representing a computer program that can cause a computer to function in the particular fashion described herein.

Those skilled in the art will understand that the systems and methods of the present invention may be applied to a variety of systems, including IBM-compatible personal computers running MS-DOS or Microsoft Windows.

The computer may have internal components linked to external components. The internal components may include a processor element interconnected with a main memory. The computer system can be an Intel Pentium[®]-based processor of 200 MHz or greater clock rate and with 32 MB or more of main memory. The external component may comprise a mass storage, which can be one or more hard disks (which are typically packaged together with the processor and memory). Such hard disks are typically of 1 GB or greater storage capacity. Other external components include a user interface device, which can be a monitor, together with an inputting device, which can be a "mouse", or other graphic input devices, and/or a keyboard. A printing device can also be attached to the computer.
20

25 Typically, the computer system is also linked to a network link, which can be part of an Ethernet link to other local computer systems, remote computer systems, or wide area communication networks, such as the Internet. This network link allows the computer system to share data and processing tasks with other computer systems.

Loaded into memory during operation of this system are several software components,
30 which are both standard in the art and special to the instant invention. These software components collectively cause the computer system to function according to the methods of this

invention. These software components are typically stored on a mass storage. A software component represents the operating system, which is responsible for managing the computer system and its network interconnections. This operating system can be, for example, of the Microsoft Windows' family, such as Windows 95, Windows 98, or Windows NT. A software component represents common languages and functions conveniently present on this system to assist programs implementing the methods specific to this invention. Many high or low level computer languages can be used to program the analytic methods of this invention. Instructions can be interpreted during run-time or compiled. Preferred languages include C/C++, and JAVA[®]. Most preferably, the methods of this invention are programmed in mathematical software packages which allow symbolic entry of equations and high-level specification of processing, including algorithms to be used, thereby freeing a user of the need to procedurally program individual equations or algorithms. Such packages include Matlab from Mathworks (Natick, Mass.), Mathematica from Wolfram Research (Champaign, Ill.), or S-Plus from Math Soft (Cambridge, Mass.). Accordingly, a software component represents the analytic methods of this invention as programmed in a procedural language or symbolic package. In a preferred embodiment, the computer system also contains a database comprising values representing levels of expression of one or more genes which are up- or down-regulated in R.A. The database may contain one or more expression profiles of genes which are up- or down-regulated in R.A. in different cells.

In an exemplary implementation, to practice the methods of the present invention, a user first loads expression data into the computer system. These data can be directly entered by the user from a monitor and keyboard, or from other computer systems linked by a network connection, or on removable storage media such as a CD-ROM or floppy disk or through the network. Next the user causes execution of expression profile analysis software which performs the steps of comparing and, e.g., clustering co-varying genes into groups of genes.

In another exemplary implementation, expression profiles are compared using a method described in U.S. Patent No. 6,203,987. A user first loads expression profile data into the computer system. Geneset profile definitions are loaded into the memory from the storage media or from a remote computer, preferably from a dynamic geneset database system, through the network. Next the user causes execution of projection software which performs the steps of

converting expression profile to projected expression profiles. The projected expression profiles are then displayed.

In yet another exemplary implementation, a user first loads a projected profile into the memory. The user then causes the loading of a reference profile into the memory. Next, the user causes the execution of comparison software which performs the steps of objectively comparing the profiles.

3. Exemplary diagnostic and prognostic compositions and devices of the invention

Any composition and device (e.g., an array) used in the above-described methods are within the scope of the invention.

In one embodiment, the invention provides a composition comprising a plurality of detection agents for detecting expression of genes which are up- or down-regulated in R.A. In a preferred embodiment, the composition comprises at least 2, preferably at least 3, 5, 10, 20, 50, or 100 different detection agents. A detection agent can be a nucleic acid probe, e.g., DNA or RNA, or it can be a polypeptide, e.g., as antibody that binds to the polypeptide encoded by a gene characteristic of R.A. The probes can be present in equal amount or in different amounts in the solution.

A nucleic acid probe can be at least about 10 nucleotides long, preferably at least about 15, 20, 25, 30, 50, 100 nucleotides or more, and can comprise the full length gene. Preferred probes are those that hybridize specifically to genes listed in any of Tables 1-5. If the nucleic acid is short (i.e., 20 nucleotides or less), the sequence is preferably perfectly complementary to the target gene (i.e., a gene that is characteristic of R.A.), such that specific hybridization can be obtained. However, nucleic acids, even short ones that are not perfectly complementary to the target gene can also be included in a composition of the invention, e.g., for use as a negative control. Certain compositions may also comprise nucleic acids that are complementary to, and capable of detecting, an allele of a gene.

In a preferred embodiment, the invention provides nucleic acids which hybridize under high stringency conditions of 0.2 to 1 x SSC at 65 °C followed by a wash at 0.2 x SSC at 65 °C to genes which are up- or down-regulated in R.A. In another embodiment, the invention provides nucleic acids which hybridize under low stringency conditions of 6 x SSC at room temperature followed by a wash at 2 x SSC at room temperature. Other nucleic acids probes

hybridize to their target in 3 x SSC at 40 or 50 °C, followed by a wash in 1 or 2 x SSC at 20, 30, 40, 50, 60, or 65 °C.

Nucleic acids which are at least about 80%, preferably at least about 90%, even more preferably at least about 95% and most preferably at least about 98% identical to genes which are up- or down-regulated in R.A. or cDNAs thereof, and complements thereof, are also within the scope of the invention.

Nucleic acid probes can be obtained by, e.g., polymerase chain reaction (PCR) amplification of gene segments from genomic DNA, cDNA (e.g., by RT-PCR), or cloned sequences. PCR primers are chosen, based on the known sequence of the genes or cDNA, that result in amplification of unique fragments. Computer programs can be used in the design of primers with the required specificity and optimal amplification properties. *See, e.g.,* Oligo version 5.0 (National Biosciences). Factors which apply to the design and selection of primers for amplification are described, for example, by Rylchik, W. (1993) "Selection of Primers for Polymerase Chain Reaction," in *Methods in Molecular Biology*, Vol. 15, White B. ed., Humana Press, Totowa, N.J. Sequences can be obtained from GenBank or other public sources.

Oligonucleotides of the invention may be synthesized by standard methods known in the art, e.g. by use of an automated DNA synthesizer (such as are commercially available from Biosearch, Applied Biosystems, etc.). As examples, phosphorothioate oligonucleotides may be synthesized by the method of Stein et al. (1988, *Nucl. Acids Res.* 16: 3209), methylphosphonate oligonucleotides can be prepared by use of controlled pore glass polymer supports (Sarin et al., 1988, *Proc. Nat. Acad. Sci. U.S.A.* 85: 7448-7451), etc. In another embodiment, the oligonucleotide is a 2'-O-methylribonucleotide (Inoue et al., 1987, *Nucl. Acids Res.* 15: 6131-6148), or a chimeric RNA-DNA analog (Inoue et al., 1987, *FEBS Lett.* 215: 327-330).

"Rapid amplification of cDNA ends," or RACE, is a PCR method that can be used for amplifying cDNAs from a number of different RNAs. The cDNAs may be ligated to an oligonucleotide linker and amplified by PCR using two primers. One primer may be based on sequence from the instant nucleic acids, for which full length sequence is desired, and a second primer may comprise a sequence that hybridizes to the oligonucleotide linker to amplify the cDNA. A description of this method is reported in PCT Pub. No. WO 97/19110.

In another embodiment, the invention provides a composition comprising a plurality of agents which can detect a polypeptide encoded by a gene characteristic of R.A. An agent can be,

e.g., an antibody. Antibodies to polypeptides described herein can be obtained commercially, or they can be produced according to methods known in the art.

The probes can be attached to a solid support, such as paper, membranes, filters, chips, pins or glass slides, or any other appropriate substrate, such as those further described herein.

5 For example, probes of genes which are up- or down-regulated in R.A. can be attached covalently or non covalently to membranes for use, e.g., in dotblots, or to solids such as to create arrays, e.g., microarrays.

4. Therapeutic methods and compositions for R.A.

10 The expression profiling results described in the Examples indicated that certain genes are expressed at higher levels and certain genes are expressed at lower levels in cells of R.A. patients relative to their expression in normal counterpart cells. For example, SOCS3 (CISH3); RAGE (AGER); LST-1 (LY117); SAA 1-3; HMG-1; S100 A8, A9, and A12; SLPI; GILZ; PTPN-18; GADD-45A and B; Legumain (PRSC1); FST1; Lcn2; GPI; SpiL; and TSG-6 are over-expressed in patients relative to controls. Exemplary genes which are down-regulated include CMAK2B, PLA2G2A, GBAS and SOX15. Accordingly, reducing the expression of one or more of genes that are up-regulated and/or increasing the expression of one or more genes which are down-regulated in diseased cells may provide a method of treatment of R.A. Genes, whose normalization of expression improves R.A. can be identified according to methods known in the art, some of which are set forth below. Accordingly, the invention provides compositions for these therapeutic methods, e.g., compositions comprising isolated polypeptides encoded by a gene selected from the group consisting of SOCS3 (CISH3); RAGE (AGER); LST-1 (LY117); SAA 1-3; HMG-1; S100 A8, A9, and A12; SLPI; GILZ; PTPN-18; GADD-45A and B; Legumain (PRSC1); FST1; Lcn2; GPI; SpiL; and TSG-6; nucleic acids encoding such; plasmids, 25 vectors and host cell comprising these isolated nucleic acids; methods for making a polypeptide; and methods for identifying compounds which modulate gene expression of the genes or the activity of a polypeptide encoded by the genes.

4.1. Methods for determining whether modulation of the expression of a gene improves R.A.

30 In one embodiment, the effect of up- or down-regulating the level of expression of a gene which is down- or up-regulated, respectively, in a cell characteristic of R.A. is determined by

phenotypic analysis of the cell, in particular by determining whether the cell adopts a phenotype that is more reminiscent of that of a normal cell than that of a cell characteristic of R.A.

In another preferred embodiment, the effect on the cell is determined by measuring the level of expression of one or more genes which are up- or down-regulated in R.A., and preferably at least about 10, or at least about 100 genes characteristic of R.A. In a preferred embodiment, the level of expression of a gene is modulated, and the level of expression of at least one gene characteristic of R.A. is determined, e.g., by using a microarray having probes to the one or more genes. If the normalization of expression of the gene results in at least some normalization of the gene expression profile in the diseased cell, then normalizing the expression of the gene in a subject having R.A. is expected to improve R.A. The term "normalization of the expression of a gene in a diseased cell" refers to bringing the level of expression of that gene in the diseased cell to a level that is similar to that in the corresponding normal cell. "Normalization of the gene expression profile in a diseased cell" refers to bringing the expression profile in a diseased cell essentially to that in the corresponding non-diseased cell. If, however, the normalization of expression of the gene does not result in at least some normalization of the gene expression profile in the diseased cell, normalizing the expression of the gene in a subject having R.A. is not expected to improve R.A. In certain embodiments, the expression level of two or more genes which are up- or down-regulated in R.A. is modulated and the effect on the diseased cell is determined.

A preferred cell for use in these assays is a cell characteristic of R.A. that can be obtained from a subject and, e.g., established as a primary cell culture. The cell can be immortalized by methods known in the art, e.g., by expression of an oncogene or large T antigen of SV40. Alternatively, cell lines corresponding to such a diseased cell can be used. Examples include RAW cells and THP1 cells. However, prior to using such cell lines, it may be preferably to confirm that the gene expression profile of the cell line corresponds essentially to that of a cell characteristic of R.A.. This can be done as described in details herein.

Modulating the expression of a gene in a cell can be achieved, e.g., by contacting the cell with an agent that increases the level of expression of the gene or the activity of the polypeptide encoded by the gene. Increasing the level of a polypeptide in a cell can also be achieved by transfecting the cell, transiently or stably, with a nucleic acid encoding the polypeptide. Decreasing the expression of a gene in a cell can be achieved by inhibiting transcription or

translation of the gene or RNA, e.g., by introducing antisense nucleic acids, ribozymes or siRNAs into the cells, or by inhibiting the activity of the polypeptide encoded by the gene, e.g., by using antibodies or dominant negative mutants. These methods are further described below in the context of therapeutic methods.

5 A nucleic acid encoding a particular polypeptide can be obtained, e.g., by RT-PCR from a cell that is known to express the gene. Primers for the RT-PCR can be derived from the nucleotide sequence of the gene encoding the polypeptide. The nucleotide sequence of the gene is available, e.g., in GenBank or in the publications. GenBank Accession numbers of the genes listed in Tables 1-5 are provided in the tables. Amplified DNA can then be inserted into an
10 expression vector, according to methods known in the art and transfected into diseased cells of R.A. In a control experiment, normal counterpart cells can also be transfected. The level of expression of the polypeptide in the transfected cells can be determined, e.g., by electrophoresis and staining of the gel or by Western blot using an agent that binds the polypeptide, e.g., an antibody. The level of expression of one or more genes which are up- or down-regulated in R.A. can then be determined in the transfected cells having elevated levels of the polypeptide. In a preferred embodiment, the level of expression is determined by using a microarray. For example, RNA is extracted from the transfected cells, and used as target DNA for hybridization to a microarray, as further described herein.

15 These assays will allow the identification of genes which are up- or down-regulated in R.A. which can be used as therapeutic targets for developing therapeutics for R.A.

4.2. Therapeutic methods

4.2.1. Methods for reducing expression of gene in the cells of a patient

Genes which are up-regulated in R.A. may be used as therapeutic targets for treating R.A.
25 For example, it may be possible to treat R.A. by decreasing the level of the polypeptide in the diseased cells.

(i) *Antisense nucleic acids*

One method for decreasing the level of expression of a gene is to introduce into the cell antisense molecules which are complementary to at least a portion of the gene or RNA of the
30 gene. An “antisense” nucleic acid as used herein refers to a nucleic acid capable of hybridizing to a sequence-specific (e.g., non-poly A) portion of the target RNA, for example its translation

initiation region, by virtue of some sequence complementarity to a coding and/or non-coding region. The antisense nucleic acids of the invention can be oligonucleotides that are double-stranded or single-stranded, RNA or DNA or a modification or derivative thereof, which can be directly administered in a controllable manner to a cell or which can be produced intracellularly by transcription of exogenous, introduced sequences in controllable quantities sufficient to perturb translation of the target RNA.

Preferably, antisense nucleic acids are of at least six nucleotides and are preferably oligonucleotides (ranging from 6 to about 200 oligonucleotides). In specific aspects, the oligonucleotide is at least 10 nucleotides, at least 15 nucleotides, at least 100 nucleotides, or at least 200 nucleotides. The oligonucleotides can be DNA or RNA or chimeric mixtures or derivatives or modified versions thereof, single-stranded or double-stranded. The oligonucleotide can be modified at the base moiety, sugar moiety, or phosphate backbone. The oligonucleotide may include other appending groups such as peptides, or agents facilitating transport across the cell membrane (*see, e.g.*, Letsinger et al., 1989, Proc. Natl. Acad. Sci. U.S.A. 86: 6553-6556; Lemaitre et al., 1987, Proc. Natl. Acad. Sci. 84: 648-652; PCT Publication No. WO 88/09810, published Dec. 15, 1988), hybridization-triggered cleavage agents (*see, e.g.*, Krol et al., 1988, BioTechniques 6: 958-976) or intercalating agents (*see, e.g.*, Zon, 1988, Pharm. Res. 5: 539-549).

In a preferred aspect of the invention, an antisense oligonucleotide is provided, preferably as single-stranded DNA. The oligonucleotide may be modified at any position on its structure with constituents generally known in the art. For example, the antisense oligonucleotides may comprise at least one modified base moiety which is selected from the group including but not limited to 5-fluorouracil, 5-bromouracil, 5-chlorouracil, 5-iodouracil, hypoxanthine, xanthine, 4-acetylcytosine, 5-(carboxyhydroxymethyl) uracil, 5-carboxymethylaminomethyl-2-thiouridine, 5-carboxymethylaminomethyluracil, dihydrouracil, beta-D-galactosylqueosine, inosine, N6-isopentenyladenine, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 2-methyladenine, 2-methylguanine, 3-methylcytosine, 5-methylcytosine, N6-adenine, 7-methylguanine, 5-methylaminomethyluracil, 5-methoxyaminomethyl-2-thiouracil, beta-D-mannosylqueosine, 5'-methoxycarboxymethyluracil, 5-methoxyuracil, 2-methylthio-N6-isopentenyladenine, uracil-5-oxyacetic acid (v), wybutoxosine, pseudouracil, queosine, 2-thiocytosine, 5-methyl-2-thiouracil, 2-thiouracil, 4-thiouracil, 5-methyluracil, uracil-5-oxyacetic acid methylester, uracil-5-oxyacetic

acid (v), 5-methyl-2-thiouracil, 3-(3-amino-3-N-2-carboxypropyl)uracil, (acp3)w, and 2,6-diaminopurine.

In another embodiment, the oligonucleotide comprises at least one modified sugar moiety selected from the group including, but not limited to, arabinose, 2-fluoroarabinose, xylulose, and
5 hexose.

In yet another embodiment, the oligonucleotide comprises at least one modified phosphate backbone selected from the group consisting of a phosphorothioate, a phosphorodithioate, a phosphoramidothioate, a phosphoramidate, a phosphordiamidate, a methylphosphonate, an alkyl phosphotriester, and a formacetal or analog thereof.

10 In yet another embodiment, the oligonucleotide is a 2- α -anomeric oligonucleotide. An α -anomeric oligonucleotide forms specific double-stranded hybrids with complementary RNA in which, contrary to the usual β -units, the strands run parallel to each other (Gautier et al., 1987, Nucl. Acids Res. 15:6625-6641).

The oligonucleotide may be conjugated to another molecule, e.g., a peptide, hybridization triggered cross-linking agent transport agent, hybridization-triggered cleavage agent, etc. An antisense molecule can be a "peptide nucleic acid" (PNA). PNA refers to an antisense molecule or anti-gene agent which comprises an oligonucleotide of at least about 5 nucleotides in length linked to a peptide backbone of amino acid residues ending in lysine. The terminal lysine confers solubility to the composition. PNAs preferentially bind complementary single stranded
20 DNA or RNA and stop transcript elongation, and may be pegylated to extend their lifespan in the cell.

The antisense nucleic acids of the invention comprise a sequence complementary to at least a portion of a target RNA species. However, absolute complementarity, although preferred, is not required. A sequence "complementary to at least a portion of an RNA," as referred to
25 herein, means a sequence having sufficient complementarity to be able to hybridize with the RNA, forming a stable duplex; in the case of double-stranded antisense nucleic acids, a single strand of the duplex DNA may thus be tested, or triplex formation may be assayed. The ability to hybridize will depend on both the degree of complementarity and the length of the antisense nucleic acid. Generally, the longer the hybridizing nucleic acid, the more base mismatches with a
30 target RNA it may contain and still form a stable duplex (or triplex, as the case may be). One skilled in the art can ascertain a tolerable degree of mismatch by use of standard procedures to

determine the melting point of the hybridized complex. The amount of antisense nucleic acid that will be effective in the inhibiting translation of the target RNA can be determined by standard assay techniques.

5 The synthesized antisense oligonucleotides can then be administered to a cell in a controlled manner. For example, the antisense oligonucleotides can be placed in the growth environment of the cell at controlled levels where they may be taken up by the cell. The uptake of the antisense oligonucleotides can be assisted by use of methods well known in the art.

10 In an alternative embodiment, the antisense nucleic acids of the invention are controllably expressed intracellularly by transcription from an exogenous sequence. For example, a vector can be introduced *in vivo* such that it is taken up by a cell, within which cell the vector or a portion thereof is transcribed, producing an antisense nucleic acid (RNA) of the invention. Such a vector would contain a sequence encoding the antisense nucleic acid. Such a vector can remain episomal or become chromosomally integrated, as long as it can be transcribed to produce the desired antisense RNA. Such vectors can be constructed by recombinant DNA technology methods standard in the art. Vectors can be plasmid, viral, or others known in the art, used for replication and expression in mammalian cells. Expression of the sequences encoding the antisense RNAs can be by any promoter known in the art to act in a cell of interest. Such promoters can be inducible or constitutive. Most preferably, promoters are controllable or inducible by the administration of an exogenous moiety in order to achieve controlled expression of the antisense oligonucleotide. Such controllable promoters include the Tet promoter. Other usable promoters for mammalian cells include, but are not limited to: the SV40 early promoter region (Bernoist and Chambon, 1981, Nature 290: 304-310), the promoter contained in the 3' long terminal repeat of Rous sarcoma virus (Yamamoto et al., 1980, Cell 22: 787-797), the herpes thymidine kinase promoter (Wagner et al., 1981, Proc. Natl. Acad. Sci. U.S.A. 78: 1441-1445), the regulatory sequences of the metallothionein gene (Brinster et al., 1982, Nature 296: 39-42), etc.

25 Antisense therapy for a variety of cancers is in clinical phase and has been discussed extensively in the literature. Reed reviewed antisense therapy directed at the Bcl-2 gene in tumors; gene transfer-mediated overexpression of Bcl-2 in tumor cell lines conferred resistance to many types of cancer drugs. (Reed, J.C., *N.C.I.* (1997) 89:988-990). The potential for clinical development of antisense inhibitors of *ras* is discussed by Cowser, L.M., *Anti-Cancer Drug*

Design (1997) 12:359-371. Additional important antisense targets include leukemia (Geurtz, A.M., *Anti-Cancer Drug Design* (1997) 12:341-358); human C-ref kinase (Monia, B.P., *Anti-Cancer Drug Design* (1997) 12:327-339); and protein kinase C (McGraw *et al.*, *Anti-Cancer Drug Design* (1997) 12:315-326).

5 (ii) *Ribozymes*

In another embodiment, the level of a particular mRNA or polypeptide in a cell is reduced by introduction of a ribozyme into the cell or nucleic acid encoding such. Ribozyme molecules designed to catalytically cleave mRNA transcripts can also be introduced into, or expressed, in cells to inhibit expression of the gene (*see, e.g.*, Sarver *et al.*, 1990, *Science* 247:1222-1225 and U.S. Patent No. 5,093,246). One commonly used ribozyme motif is the hammerhead, for which the substrate sequence requirements are minimal. Design of the hammerhead ribozyme is disclosed in Usman *et al.*, *Current Opin. Struct. Biol.* (1996) 6:527-533. Usman also discusses the therapeutic uses of ribozymes. Ribozymes can also be prepared and used as described in Long *et al.*, *FASEB J.* (1993) 7:25; Symons, *Ann. Rev. Biochem.* (1992) 61:641; Perrotta *et al.*, *Biochem.* (1992) 31:16-17; Ojwang *et al.*, *Proc. Natl. Acad. Sci. (USA)* (1992) 89:10802-10806; and U.S. Patent No. 5,254,678. Ribozyme cleavage of HIV-I RNA is described in U.S. Patent No. 5,144,019; methods of cleaving RNA using ribozymes is described in U.S. Patent No. 5,116,742; and methods for increasing the specificity of ribozymes are described in U.S. Patent No. 5,225,337 and Koizumi *et al.*, *Nucleic Acid Res.* (1989) 17:7059-7071. Preparation and use of ribozyme fragments in a hammerhead structure are also described by Koizumi *et al.*, *Nucleic Acids Res.* (1989) 17:7059-7071. Preparation and use of ribozyme fragments in a hairpin structure are described by Chowrira and Burke, *Nucleic Acids Res.* (1992) 20:2835. Ribozymes can also be made by rolling transcription as described in Daubendiek and Kool, *Nat. Biotechnol.* (1997) 15(3):273-277.

25 (iii) *siRNAs*

Another method for decreasing or blocking gene expression is by introducing double stranded small interfering RNAs (siRNAs), which mediate sequence specific mRNA degradation. RNA interference (RNAi) is the process of sequence-specific, post-transcriptional gene silencing in animals and plants, initiated by double-stranded RNA (dsRNA) that is homologous in sequence to the silenced gene. *In vivo*, long dsRNA is cleaved by ribonuclease III to generate 21- and 22-nucleotide siRNAs. It has been shown that 21-nucleotide siRNA

duplexes specifically suppress expression of endogenous and heterologous genes in different mammalian cell lines, including human embryonic kidney (293) and HeLa cells (Elbashir et al. Nature 2001 ;411(6836):494-8).

(iv) *Triplex formation*

5 Gene expression can be reduced by targeting deoxyribonucleotide sequences complementary to the regulatory region of the target gene (i.e., the gene promoter and/or enhancers) to form triple helical structures that prevent transcription of the gene in target cells in the body. (See generally, Helene, C. 1991, Anticancer Drug Des., 6(6):569-84; Helene, C., et al., 1992, Ann. N.Y. Acad. Sci., 660:27-36; and Maher, L.J., 1992, Bioassays 14(12):807-15).

10 (v) *Aptamers*

In a further embodiment, RNA aptamers can be introduced into or expressed in a cell. RNA aptamers are specific RNA ligands for proteins, such as for Tat and Rev RNA (Good et al., 1997, Gene Therapy 4: 45-54) that can specifically inhibit their translation.

(vi) *Dominant negative mutants*

15 Another method of decreasing the biological activity of a polypeptide is by introducing into the cell a dominant negative mutant. A dominant negative mutant polypeptide will interact with a molecule with which the polypeptide normally interacts, thereby competing for the molecule, but since it is biologically inactive, it will inhibit the biological activity of the polypeptide. A dominant negative mutant can be created by mutating the substrate-binding domain, the catalytic domain, or a cellular localization domain of the polypeptide. Preferably, the mutant polypeptide will be overproduced. Point mutations are made that have such an effect. In addition, fusion of different polypeptides of various lengths to the terminus of a protein can yield dominant negative mutants. General strategies are available for making dominant negative mutants. See Herskowitz, Nature (1987) 329:219-222.

25 (vi) *Use of agents inhibiting transcription or polypeptide activity*

In another embodiment, a compound decreasing the expression of the gene of interest or the activity of the polypeptide is administered to a subject having R.A., such that the level of the polypeptide in the diseased cells decreases, and the disease is improved. Compounds may be known in the art or can be identified as further described herein.

30

4.2.2. Methods for increasing the expression of a protein in cells of a patient

Genes which are down-regulated in R.A. may be used as therapeutic targets for treating R.A. For example, it may be possible to treat R.A. by increasing the level of the polypeptide in the diseased cells.

5 (i) *Administration of a nucleic acid encoding a polypeptide of interest to a subject*

In one embodiment, a nucleic acid encoding a polypeptide of interest, or an equivalent thereof, such as a functionally active fragment of the polypeptide, is administered to a subject, such that the nucleic acid arrives at the site of the diseased cells, traverses the cell membrane and is expressed in the diseased cell.

10 A nucleic acid encoding a polypeptide of interest can be obtained as described herein, e.g., by RT-PCR, or from publicly available DNA clones. It may not be necessary to express the full length polypeptide in a cell of a subject, and a functional fragment thereof may be sufficient. Similarly, it is not necessary to express a polypeptide having an amino acid sequence that is identical to that of the wild-type polypeptide. Certain amino acid deletions, additions and substitutions are permitted, provided that the polypeptide retains most of its biological activity. For example, it is expected that polypeptides having conservative amino acid substitutions will have the same activity as the polypeptide. Polypeptides that are shorter or longer than the wild-type polypeptide or which contain from one to 20 amino acid deletions, insertions or substitutions and which have a biological activity that is essentially identical to that of the wild-type polypeptide are referred to herein as “equivalents of the polypeptide.” Equivalent polypeptides also include polypeptides having an amino acid sequence which is at least 80%, preferably at least about 90%, even more preferably at least about 95% and most preferably at least 98% identical or similar to the amino acid sequence of the wild-type polypeptide.

25 Determining which portion of the polypeptide is sufficient for improving R.A. or which polypeptides derived from the polypeptide are “equivalents” which can be used for treating R.A., can be done in *in vitro* assays. For example, expression plasmids encoding various portions of the polypeptide can be transfected into cells, e.g., diseased cells of R.A., and the effect of the expression of the portion of the polypeptide in the cells can be determined, e.g., by visual inspection of the phenotype of the cell or by obtaining the expression profile of the cell, as
30 further described herein.

Any means for the introduction of polynucleotides into mammals, human or non-human, may be adapted to the practice of this invention for the delivery of the various constructs of the invention into the intended recipient. In one embodiment of the invention, the DNA constructs are delivered to cells by transfection, i.e., by delivery of "naked" DNA or in a complex with a colloidal dispersion system. A colloidal system includes macromolecule complexes, nanocapsules, microspheres, beads, and lipid-based systems including oil-in-water emulsions, micelles, mixed micelles, and liposomes. The preferred colloidal system of this invention is a lipid-complexed or liposome-formulated DNA. In the former approach, prior to formulation of DNA, e.g., with lipid, a plasmid containing a transgene bearing the desired DNA constructs may first be experimentally optimized for expression (e.g., inclusion of an intron in the 5' untranslated region and elimination of unnecessary sequences (Felgner, et al., Ann NY Acad Sci 126-139, 1995). Formulation of DNA, e.g. with various lipid or liposome materials, may then be effected using known methods and materials and delivered to the recipient mammal. See, e.g., Canonico et al, Am J Respir Cell Mol Biol 10:24-29, 1994; Tsan et al, Am J Physiol 268; Alton et al., Nat Genet. 5:135-142, 1993 and U.S. patent No. 5,679,647 by Carson et al.

The targeting of liposomes can be classified based on anatomical and mechanistic factors. Anatomical classification is based on the level of selectivity, for example, organ-specific, cell-specific, and organelle-specific. Mechanistic targeting can be distinguished based upon whether it is passive or active. Passive targeting utilizes the natural tendency of liposomes to distribute to cells of the reticulo-endothelial system (RES) in organs, which contain sinusoidal capillaries. Active targeting, on the other hand, involves alteration of the liposome by coupling the liposome to a specific ligand such as a monoclonal antibody, sugar, glycolipid, or protein, or by changing the composition or size of the liposome in order to achieve targeting to organs and cell types other than the naturally occurring sites of localization.

The surface of the targeted delivery system may be modified in a variety of ways. In the case of a liposomal targeted delivery system, lipid groups can be incorporated into the lipid bilayer of the liposome in order to maintain the targeting ligand in stable association with the liposomal bilayer. Various linking groups can be used for joining the lipid chains to the targeting ligand. Naked DNA or DNA associated with a delivery vehicle, e.g., liposomes, can be administered to several sites in a subject (see below).

In a preferred method of the invention, the DNA constructs are delivered using viral vectors. The transgene may be incorporated into any of a variety of viral vectors useful in gene therapy, such as recombinant retroviruses, adenovirus, adeno-associated virus (AAV), and herpes simplex virus-1, or recombinant bacterial or eukaryotic plasmids. While various viral vectors may be used in the practice of this invention, AAV- and adenovirus-based approaches are of particular interest. Such vectors are generally understood to be the recombinant gene delivery system of choice for the transfer of exogenous genes *in vivo*, particularly into humans.

It is possible to limit the infection spectrum of viruses by modifying the viral packaging proteins on the surface of the viral particle (see, for example PCT publications WO93/25234, WO94/06920, and WO94/11524). For instance, strategies for the modification of the infection spectrum of viral vectors include: coupling antibodies specific for cell surface antigens to envelope protein (Roux et al., (1989) PNAS USA 86:9079-9083; Julan et al., (1992) J. Gen Virol 73:3251-3255; and Goud et al., (1983) Virology 163:251-254); or coupling cell surface ligands to the viral envelope proteins (Neda et al., (1991) J. Biol. Chem. 266:14143-14146). Coupling can be in the form of the chemical cross-linking with a protein or other variety (e.g. lactose to convert the env protein to an asialoglycoprotein), as well as by generating fusion proteins (e.g. single-chain antibody/env fusion proteins). This technique, while useful to limit or otherwise direct the infection to certain tissue types, and can also be used to convert an ecotropic vector in to an amphotropic vector.

The expression of a polypeptide of interest or equivalent thereof in cells of a patient to which a nucleic acid encoding the polypeptide was administered can be determined, e.g., by obtaining a sample of the cells of the patient and determining the level of the polypeptide in the sample, relative to a control sample. The successful administration to a patient and expression of the polypeptide or an equivalent thereof in the cells of the patient can be monitored by determining the expression of at least one gene characteristic of R.A., and preferably by determining an expression profile including most of the genes which are up- or down-regulated in R.A., as described herein.

(ii) *Administration of a polypeptide of interest to a subject*

In another embodiment, a polypeptide of interest, or an equivalent thereof, e.g., a functional fragment thereof, is administered to the subject such that it reaches the diseased cells of R.A., and traverses the cellular membrane. Polypeptides can be synthesized in prokaryotes or

eukaryotes or cells thereof and purified according to methods known in the art. For example, recombinant polypeptides can be synthesized in human cells, mouse cells, rat cells, insect cells, yeast cells, and plant cells. Polypeptides can also be synthesized in cell free extracts, e.g., reticulocyte lysates or wheat germ extracts. Purification of proteins can be done by various methods, e.g., chromatographic methods (*see, e.g.,* Robert K Scopes "Protein Purification: Principles and Practice" Third Ed. Springer-Verlag, N.Y. 1994). In one embodiment, the polypeptide is produced as a fusion polypeptide comprising an epitope tag consisting of about six consecutive histidine residues. The fusion polypeptide can then be purified on a Ni⁺⁺ column. By inserting a protease site between the tag and the polypeptide, the tag can be removed after purification of the peptide on the Ni⁺⁺ column. These methods are well known in the art and commercial vectors and affinity matrices are commercially available.

Administration of polypeptides can be done by mixing them with liposomes, as described above. The surface of the liposomes can be modified by adding molecules that will target the liposome to the desired physiological location.

In one embodiment, a polypeptide is modified so that its rate of traversing the cellular membrane is increased. For example, the polypeptide can be fused to a second peptide which promotes "transcytosis," e.g., uptake of the peptide by cells. In one embodiment, the peptide is a portion of the HIV transactivator (TAT) protein, such as the fragment corresponding to residues 37 -62 or 48-60 of TAT, portions which are rapidly taken up by cell *in vitro* (Green and Loewenstein, (1989) Cell 55:1179-1188). In another embodiment, the internalizing peptide is derived from the Drosophila antennapedia protein, or homologs thereof. The 60 amino acid long homeodomain of the homeo-protein antennapedia has been demonstrated to translocate through biological membranes and can facilitate the translocation of heterologous polypeptides to which it is coupled. Thus, polypeptides can be fused to a peptide consisting of about amino acids 42-58 of Drosophila antennapedia or shorter fragments for transcytosis. See for example Derossi et al. (1996) J Biol Chem 271:18188-18193; Derossi et al. (1994) J Biol Chem 269:10444-10450; and Perez et al. (1992) J Cell Sci 102:717-722.

(iii) *Use of agents stimulating transcription or polypeptide activity*

In another embodiment, a pharmaceutical composition comprising a compound that stimulates the level of expression of a gene of interest or the activity of the polypeptide in a cell is administered to a subject, such that the level of expression of the gene in the diseased cells is

increased or even restored, and R.A. is improving in the subject. Compounds may be known in the art or can be identified as further described herein.

4.3. Drug design and discovery of therapeutics

As described above, genes whose modulation of expression improve R.A. can be used as targets in drug design and discovery. For example, assays can be conducted to identify molecules that modulate the expression and or activity of genes which are up- or down-regulated in R.A.

In one embodiment, an agent which modulates the expression of a gene of interest is identified by contacting cells expressing the gene with test compounds, and monitoring the level of expression of the gene. Alternatively, compounds which modulate the expression of gene X can be identified by conducting assays using the promoter region of a gene and screening for compounds which modify binding of proteins to the promoter region. The nucleotide sequence of the promoter may be described in a publication or available in GenBank. Alternatively, the promoter region of the gene can be isolated, e.g., by screening a genomic library with a probe corresponding to the gene. Such methods are known in the art.

Inhibitors of the polypeptide can also be agents which bind to the polypeptide, and thereby prevent it from functioning normally, or which degrades or causes the polypeptide to be degraded. For example, such an agent can be an antibody or derivative thereof which interacts specifically with the polypeptide. Preferred antibodies are monoclonal antibodies, humanized antibodies, human antibodies, and single chain antibodies. Such antibodies can be prepared and tested as known in the art.

If a polypeptide of interest binds to another polypeptide, drugs can be developed which modulate the activity of the polypeptide by modulating its binding to the other polypeptide (referred to herein as "binding partner"). Cell-free assays can be used to identify compounds which are capable of interacting with the polypeptide or binding partner, to thereby modify the activity of the polypeptide or binding partner. Such a compound can, e.g., modify the structure of the polypeptide or binding partner and thereby effect its activity. Cell-free assays can also be used to identify compounds which modulate the interaction between the polypeptide and a binding partner. In a preferred embodiment, cell-free assays for identifying such compounds consist essentially in a reaction mixture containing the polypeptide and a test compound or a

library of test compounds in the presence or absence of a binding partner. A test compound can be, e.g., a derivative of a binding partner, e.g., a biologically inactive peptide, or a small molecule.

Accordingly, one exemplary screening assay of the present invention includes the steps of contacting the polypeptide or functional fragment thereof or a binding partner with a test compound or library of test compounds and detecting the formation of complexes. For detection purposes, the molecule can be labeled with a specific marker and the test compound or library of test compounds labeled with a different marker. Interaction of a test compound with a polypeptide or fragment thereof or binding partner can then be detected by determining the level of the two labels after an incubation step and a washing step. The presence of two labels after the washing step is indicative of an interaction.

An interaction between molecules can also be identified by using real-time BIA (Biomolecular Interaction Analysis, Pharmacia Biosensor AB) which detects surface plasmon resonance (SPR), an optical phenomenon. Detection depends on changes in the mass concentration of macromolecules at the biospecific interface, and does not require any labeling of interactants. In one embodiment, a library of test compounds can be immobilized on a sensor surface, e.g., which forms one wall of a micro-flow cell. A solution containing the polypeptide, functional fragment thereof, polypeptide analog or binding partner is then flown continuously over the sensor surface. A change in the resonance angle as shown on a signal recording, indicates that an interaction has occurred. This technique is further described, e.g., in BIA technology Handbook by Pharmacia.

Another exemplary screening assay of the present invention includes the steps of (a) forming a reaction mixture including: (i) a polypeptide of interest, (ii) a binding partner, and (iii) a test compound; and (b) detecting interaction of the polypeptide and the binding partner. The polypeptide and binding partner can be produced recombinantly, purified from a source, e.g., plasma, or chemically synthesized, as described herein. A statistically significant change (potentiation or inhibition) in the interaction of the polypeptide and binding partner in the presence of the test compound, relative to the interaction in the absence of the test compound, indicates a potential agonist (mimetic or potentiator) or antagonist (inhibitor) of the polypeptide bioactivity for the test compound. The compounds of this assay can be contacted simultaneously. Alternatively, the polypeptide can first be contacted with a test compound for an

appropriate amount of time, following which the binding partner is added to the reaction mixture. The efficacy of the compound can be assessed by generating dose response curves from data obtained using various concentrations of the test compound. Moreover, a control assay can also be performed to provide a baseline for comparison. In the control assay, isolated and purified polypeptide or binding partner is added to a composition containing the binding partner or polypeptide, and the formation of a complex is quantified in the absence of the test compound.

Complex formation between a polypeptide and a binding partner may be detected by a variety of techniques. Modulation of the formation of complexes can be quantitated using, for example, detectably labeled proteins such as radiolabeled, fluorescently labeled, or enzymatically labeled polypeptides or binding partners, by immunoassay, or by chromatographic detection.

Typically, it will be desirable to immobilize either the polypeptide or its binding partner to facilitate separation of complexes from uncomplexed forms of one or both of the proteins, as well as to accommodate automation of the assay. Binding of the polypeptide to a binding partner, can be accomplished in any vessel suitable for containing the reactants. Examples include microtitre plates, test tubes, and micro-centrifuge tubes. In one embodiment, a fusion protein can be provided which adds a domain that allows the protein to be bound to a matrix. For example, glutathione-S-transferase/polypeptide (GST/polypeptide) fusion proteins can be adsorbed onto glutathione sepharose beads (Sigma Chemical, St. Louis, MO) or glutathione derivatized microtitre plates, which are then combined with the binding partner, e.g. an ^{35}S -labeled binding partner, and the test compound, and the mixture incubated under conditions conducive to complex formation, e.g. at physiological conditions for salt and pH, though slightly more stringent conditions may be desired. Following incubation, the beads are washed to remove any unbound label, and the matrix immobilized and radiolabel determined directly (e.g. beads placed in scintillant), or in the supernatant after the complexes are subsequently dissociated. Alternatively, the complexes can be dissociated from the matrix, separated by SDS-PAGE, and the level of the polypeptide or binding partner found in the bead fraction quantitated from the gel using standard electrophoretic techniques such as described in the appended examples.

Other techniques for immobilizing proteins on matrices are also available for use in the subject assay. For instance, either the polypeptide or its cognate binding partner can be

immobilized utilizing conjugation of biotin and streptavidin. For instance, biotinylated polypeptide molecules can be prepared from biotin-NHS (N-hydroxy-succinimide) using techniques well known in the art (e.g., biotinylation kit, Pierce Chemicals, Rockford, IL), and immobilized in the wells of streptavidin-coated 96 well plates (Pierce Chemical). Alternatively, antibodies reactive with the polypeptide can be derivatized to the wells of the plate, and the polypeptide trapped in the wells by antibody conjugation. As above, preparations of a binding partner and a test compound are incubated in the polypeptide X presenting wells of the plate, and the amount of complex trapped in the well can be quantitated. Exemplary methods for detecting such complexes, in addition to those described above for the GST-immobilized complexes, include immunodetection of complexes using antibodies reactive with the binding partner, or which are reactive with the polypeptide and compete with the binding partner; as well as enzyme-linked assays which rely on detecting an enzymatic activity associated with the binding partner, either intrinsic or extrinsic activity. In the instance of the latter, the enzyme can be chemically conjugated or provided as a fusion protein with the binding partner. To illustrate, the binding partner can be chemically cross-linked or genetically fused with horseradish peroxidase, and the amount of polypeptide trapped in the complex can be assessed with a chromogenic substrate of the enzyme, e.g. 3,3'-diamino-benzadine tetrahydrochloride or 4-chloro-1-naphthol. Likewise, a fusion protein comprising the polypeptide and glutathione-S-transferase can be provided, and complex formation quantitated by detecting the GST activity using 1-chloro-2,4-dinitrobenzene (Habig et al (1974) J Biol Chem 249:7130).

For processes that rely on immunodetection for quantitating one of the proteins trapped in the complex, antibodies against the protein can be used. Alternatively, the protein to be detected in the complex can be "epitope tagged" in the form of a fusion protein which includes, in addition to the polypeptide sequence, a second polypeptide for which antibodies are readily available (e.g. from commercial sources). For instance, the GST fusion proteins described above can also be used for quantification of binding using antibodies against the GST moiety. Other useful epitope tags include myc-epitopes (e.g., see Ellison et al. (1991) J Biol Chem 266:21150-21157) which includes a 10-residue sequence from c-myc, as well as the pFLAG system (International Biotechnologies, Inc.) or the pEZZ-protein A system (Pharmacia, NJ).

4.4. Drug design using microarrays

The invention also provides methods for designing and optimizing drugs for R.A., e.g., those which have been identified as described herein. In one embodiment, compounds are screened by comparing the expression level of one or more genes which are up- or down-regulated in R.A. in a cell characteristic of R.A. treated with a drug relative to their expression in a reference cell. In an even more preferred embodiment, the expression level of the genes is determined using microarrays, by comparing the gene expression profile of a cell treated with a test compound with the gene expression profile of a normal counterpart cell (a "reference profile"). Optionally the expression profile is also compared to that of a cell characteristic of R.A. The comparisons are preferably done by introducing the gene expression profile data of the cell treated with the drug into a computer system comprising reference gene expression profiles which are stored in a computer readable form, using appropriate algorithms. Test compounds will be screened for those which alter the level of expression of genes which are up- or down-regulated in R.A., so as to bring them to a level that is similar to that in a cell of the same type as a cell characteristic of R.A. are. Such compounds, i.e., compounds which are capable of normalizing the expression of at least about 10%, preferably at least about 20%, 50%, 70%, 80% or 90% of the genes which are up- or down-regulated in R.A., are candidate therapeutics.

The efficacy of the compounds can then be tested in additional *in vitro* assays and *in vivo*, in animal models. Animal models of R.A. include the collagen-induced arthritis mouse model (see Examples). The test compound is administered to the test animal and one or more symptoms of the disease are monitored for improvement of the condition of the animal. Expression of one or more genes which are up- or down-regulated in R.A. can also be measured before and after administration of the test compound to the animal. A normalization of the expression of one or more of these genes is indicative of the efficiency of the compound for treating R.A. in the animal.

The toxicity of the candidate therapeutic compound, such as resulting from a stress-related response, can be evaluated, e.g., by determining whether it induces the expression of genes known to be associated with a toxic response. Expression of such toxicity related genes may be determined in different cell types, preferably those that are known to express the genes. In a preferred method, microarrays are used for detecting changes in gene expression of genes known to be associated with a toxic response. Changes in gene expression may be a more

sensitive marker of human toxicity than routine preclinical safety studies. It was shown, e.g., that a drug which was found not to be toxic in laboratory animals was toxic when administered to humans. When gene profiling was studied in cells contacted with the drug, however, it was found that a gene, whose expression is known to correlate to liver toxicity, was expressed (see below).

Such microarrays will comprise genes which are modulated in response to toxicity or stress. An exemplary array that can be used for that purpose is the Affymetrix Rat Toxicology U34 array, which contains probes of the following genes: metabolism enzymes, e.g., CYP450s, acetyltransferases, and sulfotransferases; growth factors and their receptors, e.g., IGFs, interleukins, NGTs, TGFs, and VEGT; kinases and phosphatases, e.g., lipid kinases, MAFKs, and stress-activated kinases; nuclear receptors, e.g., retinoic acid, retinoid X and PPARs; transcription factors, e.g., oncogenes, STATs, NF-kB, and zinc finger proteins; apoptosis genes, e.g., Bcl-2 genes, Bad, Bax, Caspases and Fas; stress response genes, e.g., heat-shock proteins and drug transporters; membrane proteins, e.g., gap-junction proteins and selectins; and cell-cycle regulators, e.g., cyclins and cyclin-associated proteins. Other genes included in the microarrays are only known because they contain the nucleotide sequence of an EST and because they have a connection with toxicity.

In one embodiment, a drug of interest is incubated with a cell, e.g., a cell in culture, the RNA is extracted, and expression of genes is analyzed with an array containing genes which have been shown to be up- or down-regulated in response to certain toxins. The results of the hybridization are then compared to databases containing expression levels of genes in response to certain known toxins in certain organisms. For example, the GeneLogic ToxExpressTM database can be used for that purpose. The information in this database was obtained in least in part from the use of the Affymetrix GeneChip[®] rat and human probe arrays with samples treated *in vivo* or *in vitro* with known toxins. The database contains levels of expression of liver genes in response to known liver toxins. These data were obtained by treating liver samples from rats treated *in vivo* with known toxins, and comparing the level of expression of numerous genes with that in rat or human primary hepatocytes treated *in vitro* with the same toxin. Data profiles can be retrieved and analyzed with the GeneExpressTM database tools, which are designed for complex data management and analysis. As indicated on the Affymetrix (Santa Clara, CA) website, the GeneLogic, Inc. (Gaithersburg, MD) has preformed proof of concept studies

showing the changes in gene expression levels can predict toxic events that were not identified by routine preclinical safety testing. GeneLogic tested a drug that had shown no evidence of liver toxicity in rats, but that later showed toxicity in humans. The hybridization results using the Affymetrix GeneChip[®] and GeneExpress[™] tools showed that the drug caused abnormal elevations of alanine aminotransferase (ALT), which indicates liver injury, in half of the patients who had used the drug.

In one embodiment of the invention, the drug of interest is administered to an animal, such as a mouse or a rat, at different doses. As negative controls, animals are administered the vehicle alone, e.g., buffer or water. Positive controls can consist of animals treated with drugs known to be toxic. The animals can then be sacrificed at different times, e.g., at 3, 6, and 24 hours, after administration of the drug, vehicle alone or positive control drug, mRNA extracted from a sample of their liver; and the mRNA analyzed using arrays containing nucleic acids of genes which are likely to be indicative of toxicity, e.g., the Affymetrix Rat Toxicology U34 assay. The hybridization results can then be analyzed using computer programs and databases, as described above.

In addition, toxicity of a drug in a subject can be predicted based on the alleles of drug metabolizing genes that are present in a subject. Accordingly, it is known that certain enzymes, e.g., cytochrome p450 enzymes, i.e., CYP450, metabolize drugs, and thereby may render drugs which are innocuous in certain subjects, toxic in others. A commercially available array containing probes of different alleles of such drug metabolizing genes can be obtained, e.g., from Affymetrix (Santa Clara, CA), under the name of GeneChip[®] CYP450 assay.

Thus, a drug for R.A. identified as described herein can be optimized by reducing any toxicity it may have. Compounds can be derivatized *in vitro* using known chemical methods and tested for expression of toxicity related genes. The derivatized compounds must also be retested for normalization of expression levels of genes which are up- or down-regulated in R.A. For example, the derivatized compounds can be incubated with diseased cells of R.A., and the gene expression profile determined using microarrays. Thus, incubating cells with derivatized compounds and measuring gene expression levels with a microarray that contains the genes which are up- or down-regulated in R.A. and a microarray containing toxicity related genes, compounds which are effective in treating R.A. and which are not toxic can be developed. Such compounds can further be tested in animal models as described above.

In another embodiment of the invention, a drug is developed by rational drug design, i.e., it is designed or identified based on information stored in computer readable form and analyzed by algorithms. More and more databases of expression profiles are currently being established, numerous ones being publicly available. By screening such databases for the description of drugs affecting the expression of at least some of the genes which are up- or down-regulated in R.A. in a manner similar to the change in gene expression profile from a cell characteristic of R.A. to that of a normal counterpart cell, compounds can be identified which normalize gene expression in a cell characteristic of R.A. Derivatives and analogues of such compounds can then be synthesized to optimize the activity of the compound, and tested and optimized as described above.

Compounds identified by the methods described above are within the scope of the invention. Compositions comprising such compounds, in particular, compositions comprising a pharmaceutically efficient amount of the drug in a pharmaceutically acceptable carrier are also provided. Certain compositions comprise one or more active compounds for treating R.A.

The invention also provides methods for designing therapeutics for treating diseases that are different from R.A., but related thereto. Related diseases may in fact have a gene expression profile, which even though not identical to that of R.A., will show some homology, so that drugs for treating R.A. can be used for treating the related disease or for starting the research of compounds for treating the related disease. A compound for treating R.A. can be derivatized and tested as further described herein.

4.5. Exemplary therapeutic compositions

Therapeutic compositions include the compounds described herein, e.g., in the context of therapeutic treatments of R.A. Therapeutic compositions may comprise one or more nucleic acids encoding a polypeptide characteristic of R.A., or equivalents thereof. The nucleic acids may be in expression vectors, e.g., viral vectors. Other compositions comprise one or more polypeptides characteristic of R.A., or equivalents thereof. Yet other compositions comprise nucleic acids encoding antisense RNA, or ribozymes, siRNAs or RNA aptamers. Also within the scope of the invention are compositions comprising compounds identified by the methods described herein. The compositions may comprise pharmaceutically acceptable excipients, and may be contained in a device for their administration, e.g., a syringe.

4.6. Administration of compounds and compositions of the invention

In a preferred embodiment, the invention provides a method for treating a subject having R.A., comprising administering to the subject a therapeutically effective amount of a pharmaceutical composition comprising a compound of the invention.

4.6.1. Effective Dose

Compounds of the invention refer to small molecules, polypeptides, peptide mimetics, nucleic acids or any other molecule identified as potentially useful for treating R.A.

Toxicity and therapeutic efficacy of compounds can be determined by standard pharmaceutical procedures in cell cultures or experimental animals, e.g., for determining the LD50 (The Dose Lethal To 50% Of The Population) and the ED50 (the dose therapeutically effective in 50% of the population). The dose ratio between toxic and therapeutic effects is the therapeutic index and it can be expressed as the ratio LD50/ED50. Compounds which exhibit large therapeutic indices are preferred. While compounds that exhibit toxic side effects may be used, care should be taken to design a delivery system that targets such compounds to the site of affected tissue in order to minimize potential damage to healthy cells and, thereby, reduce side effects.

Data obtained from cell culture assays and animal studies can be used in formulating a range of dosage for use in humans. The dosage of such compounds lies preferably within a range of circulating concentrations that include the ED50 with little or no toxicity. The dosage may vary within this range depending upon the dosage form employed and the route of administration utilized. For any compound used in the method of the invention, the therapeutically effective dose can be estimated initially from cell culture assays. A dose may be formulated in animal models to achieve a circulating plasma concentration range that includes the IC50 (i.e., the concentration of the test compound which achieves a half-maximal inhibition of symptoms) as determined in cell culture. Such information can be used to more accurately determine useful doses in humans. Levels in plasma may be measured, for example, by high performance liquid chromatography.

4.6.2. Formulation

Pharmaceutical compositions for use in accordance with the present invention may be formulated in conventional manner using one or more physiologically acceptable carriers or excipients. Thus, the compounds and their physiologically acceptable salts and solvates may be formulated for administration by, for example, injection, inhalation or insufflation (either through the mouth or the nose) or oral, buccal, parenteral or rectal administration. In one embodiment, the compound is administered locally, at the site where the diseased cells are present, i.e., in the blood or in a joint.

The compounds of the invention can be formulated for a variety of loads of administration, including systemic and topical or localized administration. Techniques and formulations generally may be found in Remington's Pharmaceutical Sciences, Meade Publishing Co., Easton, PA. For systemic administration, injection is preferred, including intramuscular, intravenous, intraperitoneal, and subcutaneous. For injection, the compounds of the invention can be formulated in liquid solutions, preferably in physiologically compatible buffers such as Hank's solution or Ringer's solution. In addition, the compounds may be formulated in solid form and redissolved or suspended immediately prior to use. Lyophilized forms are also included.

For oral administration, the pharmaceutical compositions may take the form of, for example, tablets, lozanges, or capsules prepared by conventional means with pharmaceutically acceptable excipients such as binding agents (e.g., pregelatinised maize starch, polyvinylpyrrolidone or hydroxypropyl methylcellulose); fillers (e.g., lactose, microcrystalline cellulose or calcium hydrogen phosphate); lubricants (e.g., magnesium stearate, talc or silica); disintegrants (e.g., potato starch or sodium starch glycolate); or wetting agents (e.g., sodium lauryl sulphate). The tablets may be coated by methods well known in the art. Liquid preparations for oral administration may take the form of, for example, solutions, syrups or suspensions, or they may be presented as a dry product for constitution with water or other suitable vehicle before use. Such liquid preparations may be prepared by conventional means with pharmaceutically acceptable additives such as suspending agents (e.g., sorbitol syrup, cellulose derivatives or hydrogenated edible fats); emulsifying agents (e.g., lecithin or acacia); non-aqueous vehicles (e.g., ationd oil, oily esters, ethyl alcohol or fractionated vegetable oils); and preservatives (e.g., methyl or propyl-p-hydroxybenzoates or sorbic acid). The preparations

may also contain buffer salts, flavoring, coloring and sweetening agents as appropriate. Preparations for oral administration may be suitably formulated to give controlled release of the active compound.

For administration by inhalation, the compounds for use according to the present invention are conveniently delivered in the form of an aerosol spray presentation from pressurized packs or a nebuliser, with the use of a suitable propellant, e.g., dichlorodifluoromethane, trichlorofluoromethane, dichlorotetrafluoroethane, carbon dioxide or other suitable gas. In the case of a pressurized aerosol the dosage unit may be determined by providing a valve to deliver a metered amount. Capsules and cartridges of e.g., gelatin for use in an inhaler or insufflator may be formulated containing a powder mix of the compound and a suitable powder base such as lactose or starch.

The compounds may be formulated for parenteral administration by injection, e.g., by bolus injection or continuous infusion. Formulations for injection may be presented in unit dosage form, e.g., in ampoules or in multi-dose containers, with an added preservative. The compositions may take such forms as suspensions, solutions or emulsions in oily or aqueous vehicles, and may contain formulatory agents such as suspending, stabilizing and/or dispersing agents. Alternatively, the active ingredient may be in powder form for constitution with a suitable vehicle, e.g., sterile pyrogen-free water, before use.

The compounds may also be formulated in rectal compositions such as suppositories or retention enemas, e.g., containing conventional suppository bases such as cocoa butter or other glycerides.

In addition to the formulations described previously, the compounds may also be formulated as a depot preparation. Such long acting formulations may be administered by implantation (for example subcutaneously or intramuscularly) or by intramuscular injection. Thus, for example, the compounds may be formulated with suitable polymeric or hydrophobic materials (for example as an emulsion in an acceptable oil) or ion exchange resins, or as sparingly soluble derivatives, for example, as a sparingly soluble salt.

Administration, e.g., systemic administration, can also be by transmucosal or transdermal means. For transmucosal or transdermal administration, penetrants appropriate to the barrier to be permeated are used in the formulation. Such penetrants are generally known in the art, and include, for example, for transmucosal administration bile salts and fusidic acid derivatives. In

addition, detergents may be used to facilitate permeation. Transmucosal administration may be through nasal sprays or using suppositories. For topical administration, the compounds of the invention can be formulated into ointments, salves, gels, or creams as generally known in the art. A wash solution can be used locally to treat an injury or inflammation to accelerate healing.

5 In clinical settings, a gene delivery system for a gene of interest can be introduced into a patient by any of a number of methods, each of which is familiar in the art. For instance, a pharmaceutical preparation of the gene delivery system can be introduced systemically, e.g., by intravenous injection, and specific transduction of the protein in the target cells occurs predominantly from specificity of transfection provided by the gene delivery vehicle, cell-type or
10 tissue-type expression due to the transcriptional regulatory sequences controlling expression of the receptor gene, or a combination thereof. In other embodiments, initial delivery of the recombinant gene is more limited with introduction into the subject or animal being quite localized. For example, the gene delivery vehicle can be introduced by catheter (*see* U.S. Patent 5,328,470) or by stereotactic injection (e.g., Chen et al. (1994) PNAS 91: 3054-3057). A nucleic acid, such as one encoding a polypeptide of interest or homologue thereof can be delivered in a gene therapy construct by electroporation using techniques described, for example, by Dev et al. ((1994) Cancer Treat Rev 20:105-115). Gene therapy can be conducted *in vivo* or *ex vivo*.

The pharmaceutical preparation of the gene therapy construct or compound of the invention can consist essentially of the gene delivery system in an acceptable diluent, or can
20 comprise a slow release matrix in which the gene delivery vehicle or compound is imbedded. Alternatively, where the complete gene delivery system can be produced intact from recombinant cells, e.g., retroviral vectors, the pharmaceutical preparation can comprise one or more cells which produce the gene delivery system.

The compositions may, if desired, be presented in a pack or dispenser device which may
25 contain one or more unit dosage forms containing the active ingredient. The pack may for example comprise metal or plastic foil, such as a blister pack. The pack or dispenser device may be accompanied by instructions for administration.

4. Exemplary kits

30 The invention further provides kits for determining the expression level of genes characteristic of disease B. The kits may be useful for identifying subjects that are predisposed

to developing R.A. or who have R.A., as well as for identifying and validating therapeutics for R.A. In one embodiment, the kit comprises a computer readable medium on which is stored one or more gene expression profiles of diseased cells of R.A., or at least values representing levels of expression of one or more genes which are up- or down-regulated in R.A. in a diseased cell.

5 The computer readable medium can also comprise gene expression profiles of counterpart normal cells, diseased cells treated with a drug, and any other gene expression profile described herein. The kit can comprise expression profile analysis software capable of being loaded into the memory of a computer system.

10 A kit can comprise a microarray comprising probes of genes which are up- or down-regulated in R.A. A kit can comprise one or more probes or primers for detecting the expression level of one or more genes which are up- or down-regulated in R.A. and/or a solid support on which probes attached and which can be used for detecting expression of one or more genes which are up- or down-regulated in R.A. in a sample. A kit may further comprise nucleic acid controls, buffers, and instructions for use.

15 Other kits provide compositions for treating R.A. For example, a kit can also comprise one or more nucleic acids corresponding to one or more genes which are up- or down-regulated in R.A., e.g., for use in treating a patient having R.A. The nucleic acids can be included in a plasmid or a vector, e.g., a viral vector. Other kits comprise a polypeptide encoded by a gene characteristic of R.A. or an antibody to a polypeptide. Yet other kits comprise compounds identified herein as agonists or antagonists of genes which are up- or down-regulated in R.A. 20 The compositions may be pharmaceutical compositions comprising a pharmaceutically acceptable excipient.

25 The present invention is further illustrated by the following examples which should not be construed as limiting in any way. The contents of all cited references including literature references, issued patents, published and non published patent applications as cited throughout this application are hereby expressly incorporated by reference.

30 The practice of the present invention will employ, unless otherwise indicated, conventional techniques of cell biology, cell culture, molecular biology, transgenic biology, microbiology, recombinant DNA, and immunology, which are within the skill of the art. Such techniques are explained fully in the literature. (See, for example, *Molecular Cloning A*

Laboratory Manual, 2nd Ed., ed. by Sambrook, Fritsch and Maniatis (Cold Spring Harbor
 Laboratory Press: 1989); *DNA Cloning*, Volumes I and II (D. N. Glover ed., 1985);
 Oligonucleotide Synthesis (M. J. Gait ed., 1984); Mullis et al. U.S. Patent No: 4,683,195;
Nucleic Acid Hybridization (B. D. Hames & S. J. Higgins eds. 1984); *Transcription And*
 5 *Translation* (B. D. Hames & S. J. Higgins eds. 1984); (R. I. Freshney, Alan R. Liss, Inc., 1987);
Immobilized Cells And Enzymes (IRL Press, 1986); B. Perbal, *A Practical Guide To Molecular*
Cloning (1984); the treatise, *Methods In Enzymology* (Academic Press, Inc., N.Y.); *Gene*
Transfer Vectors For Mammalian Cells (J. H. Miller and M. P. Calos eds., 1987, Cold Spring
 Harbor Laboratory); , Vols. 154 and 155 (Wu et al. eds.), *Immunochemical Methods In Cell And*
 10 *Molecular Biology* (Mayer and Walker, eds., Academic Press, London, 1987); *Handbook Of*
Experimental Immunology, Volumes I-IV (D. M. Weir and C. C. Blackwell, eds., 1986) (Cold
 Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1986).

Examples

Example 1: Identification of genes that are up- or down-regulated in patients having rheumatoid arthritis

This Example describes the identification of several genes which are up- or
 downregulated in peripheral blood mononuclear cells (PBMCs) of subjects having rheumatoid
 arthritis (R.A.) relative to expression in PBMCs of normal subjects.

20 PBMCs were isolated from 9 patients with R.A. and 13 normal volunteers as follows.
 Eight mls of blood were drawn into a CPT Vacutainer tube which was inverted several times.
 The tube was centrifuged at 1500 x g (2700 rpm) in a swinging bucket rotor at room temperature.
 The serum was removed and PBMCs were transferred to a 15 ml conical centrifuge tube. The
 cells were washed with the addition of phosphate buffered saline (PBS) and centrifuged at 450g
 25 (1200 rpm) for 5 minutes. The supernatant was discarded and the wash procedure was repeated
 once more. After removal of the supernatant, total RNA was isolated with the use of the RNeasy
 minikit, (Qiagen,Hidden,Germany) according to the manufacturers procedure.

RNA was analyzed on oligonucleotide arrays composed of 6,800 and 12,000 human
 genes (Affymetrix Hu6800 and HgU95A chip sets, respectively), as follows.

30 Target nucleic acid for hybridization was prepared as follows. Total RNA was prepared
 for hybridization by denaturing 5 µg of total RNA from PBMC's for 10 minutes at 70 °C with

100 pM T7/T24-tagged oligo-dt primer (synthesized at Genetics Institute, Cambridge, MA), and cooled on ice. First strand cDNA synthesis was performed under the following buffer conditions: 1x first strand buffer (Invitrogen Life Technologies, Carlsbad, CA.), 10 mM DTT(GIBCO/Invitrogen), 500 μ M of each dNTP (Invitrogen Life Technologies), 400 units of
5 Superscript RT II (Invitrogen Life Technologies) and 40 units RNase inhibitor (Ambion,Austin,TX.). The reaction proceeded at 47 $^{\circ}$ C for 1 hour. Second strand cDNA was synthesized with the addition of the following reagents at the final concentrations listed: 1X second strand buffer (Invitrogen Life Technologies), an additional 200 μ M of each dNTP (Invitrogen Life Technologies), 40 units of *E. coli* DNA polymerase I (Invitrogen Life
10 Technologies), 2 units *E. coli* RNaseH (Invitrogen Life Technologies), and 10 units of *E. coli* DNA ligase. The reaction proceeded at 15.8 $^{\circ}$ C for 2 hours and during the last five minutes of this reaction 6 units of T4 DNA polymerase (New England Biolabs, Beverly, MA) was added. The resulting double stranded cDNA was purified with the use of BioMag carboxyl terminated particles as follows: 0.2 mg of BioMag particles (Polysciences Inc., Warrington, PA) were equilibrated by washing three times with 0.5M EDTA and resuspended at a concentration of 22.2 mg/ml in 0.5M EDTA. The double stranded cDNA reaction was diluted to a final concentration of 10% PEG/1.25M NaCl and the bead suspension was added to a final bead concentration of 0.614 mg/ml. The reaction was incubated at room temperature for 10 minutes. The cDNA/ bead complexes were washed with 300 μ l of 70% ethanol, the ethanol was removed and the tubes were
15 allowed to air dry. The cDNA was eluted with the addition of 20 μ l of 10 mM Tris-acetate, pH 7.8, incubated for 2-5 minutes and the cDNA containing supernatant was removed.

10 μ l of purified double stranded cDNA was then added to an *in vitro* transcription (IVT) solution which contained, 1 x IVT buffer (Ambion, Austin,TX) 5,000 units T7 RNA polymerase (Epicentre Technologies, Madison,WI), 3mM GTP, 1.5mM ATP, 1.2mM CTP and 1.2 mM UTP
25 (Amersham/Pharmacia,), 0.4 mM each bio-16 UTP and bio-11 CTP (Enzo Diagnostics, Farmingdale, NY), and 80 units RNase inhibitor (Ambion, Austin, TX). The reaction proceeded at 37 $^{\circ}$ C for 16 hours. Labeled RNA was purified with the use of an RNeasy (Qiagen). The RNA yield was quantitated by measuring absorbance at 260nm.

Array Hybridization and Detection of Fluorescence was performed as follows. 12 μ g of
30 IVT was fragmented in 40 mM Tris-actetate, pH 8.0, 100 mM potassium acetate, and 30 mM magnesium acetate for 35 minutes at 94 $^{\circ}$ C. The fragmented, labeled RNA probes were diluted

in hybridization buffer at a final composition of 1 x 2-N-Morpholinoethanesulfonic acid (MES (buffer (pH 6.5), 50 pM Bio948 (control biotinylated oligo that hybridizes to landmark features on the probe array (Genetics Institute,Cambridge,MA)), 100 µg/ml herring sperm DNA (Promega, Madison,WI), 500 µg/ml acetylated BSA (Invitrogen Life Technologies) and 1µl/µg standard curve reagent (Proprietary reagent supplied by Gene Logic,Gaithersburg,MD). This hybridization solution was pre-hybridized with two glass beads (Fisher Scientific, Pittsburgh, PA) at 45 °C overnight. The hybridization solution was removed to a clean tube, heated for 1-2 min at 95 °C and microcentrifuged on high for 2 minutes to pellet insoluble debris. Affymetrix oligonucleotide array cartridges (human 6800 array P/N900183 and human U95A (Affymetrix, Santa Clara, CA)) were pre-wet with non-stringent wash buffer (0.9M NaCl, 60mM sodium phosphate, 6 mM EDTA and 0.01% Tween20) and incubated at 45 °C with rotation for 5-10 minutes. Buffer was removed from the Affymetrix cartridges and the arrays were hybridized with 180 µl of the hybridization solution at 45 °C rotating at 45-60 rpm overnight. After overnight incubation, the hybridization solutions were removed and the cartridges were filled with non-stringent wash buffer. The array cartridges were washed using an Affymetrix fluidics station according with 10 cycles of 2 mixes/cycle non-stringent wash buffer at 25 °C followed by 4 cycles of 15mixes/cycle stringent wash buffer (100mM MES, 0.1M Na⁺, 0.01% Tween20 and 0.005% antifoam). The probe array was then first stained for 10 minutes at 25°C in SAPE solution (100mM MES, 1M Na⁺, 0.05%Tween20, 0.005% antifoam, 2mg/ml acetylated BSA (Invitrogen Life Technologies) and 10ug/ml R phycoerythrin streptavidin (Molecular Probes, Eugene, OR)). After first staining, the probe array was washed for 10 cycles of 4 mixes/cycle with non-stringent wash buffer at 25 °C. The probe array was then stained for 10 minutes at 25 °C in antibody solution (100mM MES, 1M Na⁺, 0.05% Tween20, 0.005% antifoam, 2 mg/ml acetylated BSA (Invitrogen Life Technologies), 100µg/ml Goat IgG (SIGMA,St. Louis, MO) and 3µg/ml biotinylated anti-streptavidin antibody(goat) (Vector Laboratories). Following the second stain, the probe array is stained again for an additional 10 minutes at 25 °C in SAPE solution. Finally, the probe array is washed for 15 cycles of 4 mixes/cycle with non-stringent wash buffer at 30 °C. Arrays were scanned using an Affymetrix gene chip scanner (Affymetrix, Santa Clara, CA). The scanner contains a scanning confocal microscope and uses an argon ion laser for the excitation source and emission is detected by a photomultiplier tube at 530 nm bandpass filter (fluorescein 0 or 560 longpass filter (phycoerythrin).

Data analysis was performed using GENECHIP 3.0 or 4.0 software with normalizing/scaling to internal controls. For each patient, two parameters were used to filter the data: 1) "Absolute Decision," which indicates the presence (P) or absence (A) of RNA of a gene within a given RNA sample; 2) "Frequency," which measures the number of copies of a given RNA within a RNA sample, and this value is expressed as Copies per million transcripts. If a gene was called "Absent," its frequency was not used to calculate the average frequency of the gene. If a gene was called "Absent" for more than four patients in the Hu6800 data; more than two patients in the HgU95A data, or more than six normals, no average frequency was calculated. Genes that had average frequencies for normal volunteers only were tagged "Normal" while those that had average frequencies for patients only were tagged "Disease." The fold change in gene expression was calculated by dividing the average gene frequency of the patients by that of the normals. Genes selected for analysis met the following criteria: 1) a fold change greater than 1.95 or less than -1.95 and 2) those genes tagged as either "Normal" or "Disease."

The results are set forth in Tables 1 and 2, which are attached at the end of the written description as pages 1-66 and 1-5, respectively, and specifically incorporated by reference herein. The data in one of the columns indicates the average frequency ("Avg FC") in the patients divided by the average frequency in non-diseased subjects. An increased expression in patients relative to normals is indicated by the absence of a sign in front of the "Avg FC RA/normal" number, whereas a decrease in expression in patients relative to subjects is indicated by the presence of a "-" sign in front of the number (negative values are listed at the end of the Tables). "#DIV/0!" indicates an infinite number, resulting from expression levels in normals that are undetectable. Accession numbers, Affymetrix identifiers ("qualifiers") and gene name are depicted in the Tables. The sequence of most genes are available on GenBank. Genes whose Accession Number is characterized by "HT-...." or "HG-...." are available in the TIGR database on the internet. Genes that seemed of particular significance are highlighted or marked with a star.

Table 1 shows genes identified using the Hu6800 Affymetrix chip. Table 2 shows the genes of the Hu6800 chip that gave a positive signal and that encode a kinase or a phosphatase. Table 3 shows the genes that represent 1.95 fold or greater change in patients compared to normal, identified using the U95 chip sets. Table 3 is attached at the end of the written

description as pages 1-36, and is specifically incorporated by reference herein. The Tables indicate the chromosomal localization of the genes.

Interestingly, numerous genes which are up-regulated in R.A. patients are located on human chromosome 6 in a region (6p21.3) that contains the genes of the major histocompatibility complex (MHC) and tumor necrosis factor (TNF), suggesting that these genes may be of importance in R.A. Other genes of interest are kinases and phosphatases. Yet other genes which are of interest are those that are up- or down-regulated by a factor of at least two and those in which the ratio of induction could not be determined since no detectable signal was obtained in the normal controls (genes indicated as “#DIV/0!”). Other genes of interest are those that are highlighted or marked with a star in the Tables.

Example 2: Identification of genes which are up- or down-regulated in an animal model of rheumatoid arthritis

This example describes the identification of several genes which are up- or down-regulated in mice having collagen induced arthritis (CIA) relative to normal mice. Gene expression was measured in paws of mice; PBMCs and in synovium.

CIA is an accepted animal model for rheumatoid arthritis. The disease was induced as follows in mice. Male DBA/1 (Jackson Laboratories, Bar Harbor, Maine) mice were used for all experiments. Arthritis was induced with the use of either chicken collagen type II (Sigma, St.Louis, MO) or bovine collagen type II (Chondrex, Redmond, WA). Chicken collagen was dissolved in 0.01 M acetic acid and emulsified with an equal volume of Complete Freund's adjuvant (CFA; Difco Labs, Detroit, MI) containing 1mg/ml *Mycobacterium tuberculosis* (strain H37RA). 200 µg of chicken collagen was intradermally injected in the base of the tail on day 0. On day 21, mice were injected intraperitoneally with a PBS solution containing 100 µg of chicken collagen II. Bovine collagen type II (Chondrex, Redmond, WA) was dissolved in 0.1 M acetic acid and emulsified in an equal volume of CFA (Sigma) containing 1mg/ml *Mycobacterium tuberculosis* (strain H37RA). 200 µg of bovine collagen was injected subcutaneously in the base of the tail on day 0. On day 21, mice were injected subcutaneously, in the base of the tail, with a solution containing 200 µg of bovine collagen in 0.1 M acetic acid that had been mixed with an equal volume of Incomplete Freund's adjuvant (Sigma). Naïve animals received the same sets of injections, minus collagen. Mice were monitored at least three

times a week for disease progression. Individual limbs were assigned a clinical score based on the index: 0 = normal; P = prearthritic, characterized by focal erythema on the tips of digits.; 1 = visible erythema accompanied by 1-2 swollen digits.; 2 = pronounced erythema, characterized by paw swelling and/or multi digit swelling.; 3 = massive swelling extending into ankle or wrist joint.; 4 = difficulty in use of limb or joint rigidity. The sum of all limb scores for any given mouse could yield a maximum total body score of 16.

At various stages of disease, animals were euthanized and tissues were harvested. In one series of examples, at least two paws from each animal were flash frozen in liquid nitrogen for RNA analyses. Frozen mouse paws were pulverized to a fine powder with the use of a mortar and pestle and liquid nitrogen. RNA was purified using the Promega RNeasy Total RNA Isolation System (Promega, Madison, WI). The RNA was further purified using the RNeasy minikit. The remaining paws were fixed in 10% formalin for histology.

In another series of examples, gene expression was determined in PBMCs of mice. Blood was collected via cardiac puncture into EDTA coated collection tubes. Blood samples were pooled according to similar total body scores (normal, prearthritic, scores 1, 3, 4, 5, 6, and 7-9) into a 15 ml conical tube. The blood was diluted 1:1 with PBS that contained 2mM EDTA, and layered on an equal volume of Lympholyte-M (Cedar Lane Labs, Hornby, Ontario, Canada). The mixture was centrifuged, with no brake, for 20 minutes at 1850 rpm in a Sorvall centrifuge, (model RT 6000D). Cells at the interface were collected and added to a new tube. The cells were washed with the addition of 10 ml PBS, containing 2mM EDTA, and centrifuged at 1200 rpm for 10 minutes. The wash was repeated two times. To lyse residual red cells, cell pellets were dispersed in 2ml of cold 0.2 % NaCl and incubated on ice for 45-60 seconds. Lysis was terminated with the addition of 2 ml of 1.6 % NaCl and the cells were centrifuged at 1200 rpm for 10 minutes. PBMCs were resuspended in 5 ml of PBS, which contained 2mM EDTA, and counted. Cells were centrifuged at 1200 rpm for 10 minutes, and the supernatant discarded in preparation for RNA isolation. Total RNA was isolated from the PBMCs using the RNeasy minikit (Qiagen, Hidden, Germany).

In yet another series of examples, RNA was obtained from isolated synovium of the diseased animals. The joint synovium was dissected from diseased and control animals under a dissection scope. Tissues from five or more animals with similar disease scores were pooled and RNA was isolated using the RNeasy kit (Qiagen, Hidden, Germany).

Gene expression was analyzed on the oligonucleotide arrays Affymetrix murine 11K chip set composed of 11,000 murine genes on two chips, murine 11KsubA P/N 900188 and murine 11KsubB P/N900189.

5 Labeled target nucleic acids for hybridization to the chips were prepared as described in the previous Example with 5 µg of PBMC RNA or 7 µg of RNA from paws or synovial tissue.

10 Data analysis was performed using GENECHIP 3.0 software with normalizing/scaling to internal controls. Each experimental sample was compared to a time matched control in a two-file analysis. Next, the data were entered into the GeneSpring (Silicon Genetics, Redwood City, CA) analysis program. The data were filtered in a hierarchical fashion. First, the data were grouped according to paw scores. For each score, a list of genes that were called “Present” in all samples in a given score group and in the control was created. These lists were further refined by removing all genes that were not called either “Increasing” or “Decreasing” (defined in the program) in at least a majority of the samples in each score group. These lists were then filtered for genes that showed fold change greater than or equal to 1.95 or less than or equal to –1.95 in either all of the samples, if there were less than five samples, or in greater than 70% of the samples.

15 The results of the PBMCs are indicated in Table 4 and the results of the paw examples are indicated in Table 5. Tables 4 and 5 are attached at the end of the written description as pages 1-17 and 1-74, respectively, and are specifically incorporated by reference herein. “C” stands for control; “P” stands for prearthritic. The columns represent fold changes compared back to the normal. Accession numbers, Affymetrix identifiers (“qualifiers”) and gene name are depicted in the Tables.

20 The results show that several genes, e.g., PTPN18, HMG-1 and SLPI, that are significantly up-regulated in the mouse model, were also significantly up-regulated in human PBMCs of R.A. patients.

Example 3: Identification of cells expressing genes which are up-regulated in R.A.

This Example describes the identity of cells expressing genes which are up-regulated in R.A. by *in situ* hybridization.

Paws of CIA mice were fixed in 4% paraformaldehyde, pH 7.47, decalcified in 20% EDTA (pH 8.0) and embedded in paraffin for *in situ* hybridization according to methods known in the art.

Sense and anti-sense riboprobes for use in the *in situ* hybridization were produced by generating 2 independent PCR products, as follows. T7 RNA polymerase binding sites were incorporated into the oligonucleotides to insert T7 binding sites at either the 5' end of the PCR product for sense riboprobe or the 3' end of the PCR product for antisense riboprobe. Digoxigenin labeled probes were prepared with the use of a DIG RNA labeling mix (Roche Diagnostics, Mannheim, Germany), as described by the manufacturer, and T7 RNA polymerase (Roche Diagnostics).

The probes were obtained by PCR using the following oligonucleotide primers for each of the sense and antisense probe.

Murine SAA3 sense riboprobe:

Forward primer (with T7 site):

5'GACTGATAATACGACTCACTATAGGGCGAATGAAGCCTTCCATTGCCATCATTCTT
TGCA3' (SEQ ID NO: 1)

Reverse primer:

5'TTAGCGGCCGCTCAGTATCTTTTAGGCAGGCCAGCAGGTCGGAA3' (SEQ ID NO: 2)

The probe sequence covers the entire coding sequence, is 369 nucleotides long and has the following sequence:

atgaagcctt ccattgccat cattctttgc atcttgatcc tgggagttga cagccaaaga tgggtccagt tcatgaaaga agctgggtcaa
gggtctagag acatgtggcg agcctactct gacatgaaga aagctaactg gaaaaactca gacaaatact tccatgctcg
ggggaactat gatgctgccc ggaggggtcc cgggggagcc tgggctgcta aagtcacag cgatgccaga gaggtgttc
agaagttcac gggacatgga gcagaggact caagagctga ccagtttggc aatgagtggg gccggagtgg caaagacccc
aaccacttcc gacctgctgg cctgcctaaa agatactga (SEQ ID NO: 3)

Murine SAA3 anti-sense riboprobe:

Forward primer:

5'TTAGAATTCATGAAGCCTTCCATTGCCATCATTCTTTGCA3' (SEQ ID NO: 4)

Reverse primer (with T7 site):

5'GACTGATAATACGACTCACTATAGGGCGATCAGTATCTTTTAGGCAGGCCAGCAG
GTCGGAA3' (SEQ ID NO: 5)

The probe sequence covers the entire coding sequence, is 369 nucleotides long and has the following sequence:

tcagtatctt ttaggcaggc cagcaggtcg gaagtgggtg gggctcttgc cactccggcc ccactcattg gcaaactggt
cagctcttga gtcctctgct ccatgtcccg tgaacttctg aacagcctct ctggcatcgc tgatgacttt agcagcccag
5 gctcccccg gaccctccg ggcagcatca tagtccccc gagcatggaa gtatttgtct gagttttcc agttagcttt cttcatgtca
gagtaggctc gccacatgtc tctagaccct tgaccagctt cttcatgaa ctggacccat ctttggtgt caactcccag
gatcaagatg caaagaatga tggcaatgga aggcttcat (SEQ ID NO: 6)

Murine LST-1 sense riboprobe:

10 Forward primer (with T7 site):

100 5'GACTGATAATACGACTCACTATAGGGCGAATGTCTGATGACAATGGATCTGGTAA
CAATTGCA3'(SEQ ID NO: 7)

Reverse primer:

15 5'TTAGCGGCCGCTCAAGTGGGTGTGCTCCTGGCGATGCAGGCATA3' (SEQ ID NO: 8)

15 The probe sequence covers the entire coding sequence, has 288 nucleotides and the following sequence:

atgtctgatg acaatggatc tgtaacaat tgcacaacca atcatttctt gctctatggg agcctgggac tgggagggt
cctcctctg cttgtcatca tctgttcat ctgcctgtgc gggttcagtc agagagtga gagactggaa aggaatgcc
aggcttcagg gcaggagccc cactatgcat ctctccagca gctgccagtg tccagtagtg atatcacaga catgaaagaa
20 gacctcagca ctgactatgc ctgcatgcc aggagcacac ccacttga (SEQ ID NO: 9)

Murine LST-1 anti-sense riboprobe:

Forward primer:

5'TTAGAATTCATGTCTGATGACAATGGATCTGGTAACAATTGCA3' (SEQ ID NO: 10)

Reverse primer (with T7 site):

25 5'GACTGATAATACGACTCACTATAGGGCGATCAAGTGGGTGTGCTCCTGGCGATGC
AGGCATA3' (SEQ ID NO: 11)

The probe sequence covers the entire coding sequence, is 288nucleotides long and has the following sequence:

tcaagtgggt gtgctcctgg cgatgcaggc atagtcagt ctgaggtctt cttcatgtc tgtgatatca ctactggaca
30 ctggcagctg ctggagagat gcatagtggg gctcctgccc tgagacctgg gcattcctt ccagtctctt cactctctga

ctgaacccgc acaggcagat gaacaggatg atgacaagca ggaggaggag ccctcccagt cccaggctcc catagagcag
gaaatgattg gttgtgcaat tgttaccaga tccattgtca tcagacat (SEQ ID NO: 12)

Murine FST1 sense riboprobe:

5 Forward primer (with T7 site):

5'GACTGATAATACGACTCACTATAGGGCGAATGTGGAAACGATGGCTGGCGCTCTC
GCTG3' (SEQ ID NO: 13)

Reverse primer:

5'AGGAACAGACACAGCGATTGC3' (SEQ ID NO: 14)

10 The probe is 421 nucleotides long and has the following sequence:

atgtggaaac gatggctggc gctctcgtg gtgaccatcg ccctggcca cggcgaggag gaacctagaa gcaaatacaa
gatctgcgcc aatgtgttt gtggagctgg cagggaatgt gccgtcacag agaaggggga gcccacgtgc ctctgcattg
agcaatgcaa acctacaag aggcctgtgt gtggcagtaa tggcaagacc tacctcaacc actgtgaact tcatagagat
gcctgcctca ctggatcaa gatccagggt gattatgat ggcactgcaa agaaaagaag tctgcgagtc catctgccag
cccagttgtc tgctatcaag ctaaccgcga tgagctccga cggcgccctca tccagtggct ggaagctgag atcattccag
atggctgggt ctctaaaggc a (SEQ ID NO: 15)

Murine FST1 anti-sense riboprobe:

Forward primer:

5'GGGGATATCATGTGGAAACGATGGCTGGCGCTCTCGCTGGTGACCAT3' (SEQ ID
NO: 16)

Reverse primer (with T7 site):

5'GACTGATAATACGACTCACTATAGGGCGATGCCTTTAGAGAACCAGCCATCTGGA
ATGA3' (SEQ ID NO: 17)

The probe is 421 nucleotides long and has the following sequence:

25 tgcctttaga gaaccagcca tctggaatga tctcagcttc cagccactgg atgaggcgcc gtcggagctc atcgcggtta
gcttgatagc agacaactgg gctggcagat ggactcgcag acttcttttc ttgacgtgc ccatcataat caacctggat cttggatcca
gtgaggcagg catctctatg aagttcacag tggttgaggt aggtcttgcc attactgcca cacacaggcc tcttgtgagg
tttgattgc tcaatgcaga ggcacgtggg ctcccccttc tctgtgacgg cacattccct gccagctcca caaacacat
tggcgcagat cttggatttg cttctaggtt cctcctcgcc gtggaccagg gcgatggtea ccagcgagag cgccagccat
30 cgtttcaca t (SEQ ID NO: 18)

Murine SLPI sense riboprobe:

Forward primer (with T7 site):

5'GACTGATAATACGACTCACTATAGGGCGAATGAAGTCCTGCGGCCTTTTACCTTTC
ACGGTG3' (SEQ ID NO: 18)

5 Reverse primer:

5'AATGCGGCCGCTCACATCGGGGGCAGGCAGACTTTCCAC3' (SEQ ID NO: 19)

The probe sequence covers the entire coding sequences, is 396 nucleotides long and has the following sequence:

atgaagtct gcggcctttt acctttcacg gtgctcctg ctctgggat cctggcacc tggactgtg aaggaggcaa
10 aaatgatgct atcaaatcg gagcctgcc tgctaaaaag cctgccagc gccttaagct tgagaagcca caatgccga
ctgactggga gtgcccggga aagcagaggt gctgccaaga tgcttgcggt tccaagtgcg tgaatcctgt tccattcgc
aaaccagtgt ggaggaagcc tgggaggtgc gtcaaaactc aggcaagatg tatgatgctt aaccctcca atgtctgcca
gagggacggg cagtgtgacg gcaaatacaa gtgctgtgag ggtatatgtg ggaaagtctg cctgcccccg atgtga
(SEQ ID NO: 20)

Murine SLPI anti-sense riboprobe:

Forward primer:

5'ATTGAATTCATGAAGTCCTGCGGCCTTTTACCTTTCACGGTGC3' (SEQ ID NO: 21)

Reverse primer(with T7 site):

5'GACTGATAATACGACTCACTATAGGGCGATCACATCGGGGGCAGGCAGACTTTCC
CAC3' (SEQ ID NO: 22)

The probe sequence covers the entire coding equence, is 396 nucleotides long and has the following sequence:

tcacatcggg ggcaggcaga ctttcccaca tataacctca cagcacttgt atttgccgtc aactgcccc tccctctggc
agacattggg agggtaagc atcacatc ttgctgagt ttgacgcac ctcccaggct tctccacac tggtttgcga
25 atgggaacag gattcacgca cttggaaccg caagcatctt ggcagcacct ctgctttccc gggcactccc agtcagtacg
gcattgtggc ttctcaagct taaggcactg ggcaggcttt ttagcagggc aggctccgat ttgatagca tcatttttgc ctcttccac
agtcagggt gccaggatcc ccagagcaag gagcaccgtg aaaggtaaaa ggccgcagga cttcat (SEQ ID NO: 23)

Murine Legumain sense riboprobe:

30 Forward primer (with T7 site):

5'GACTGATAATACGACTCACTATAGGGCGA ACACCAACACCAGCCATGTC3' (SEQ ID NO: 24)

Reverse primer:

5'CTCTCAGCAGTTTCCCCAAATC3' (SEQ ID NO: 25)

5 The probe is 313 nucleotides long and has the following sequence:

acaccaacac cagccatgtc atgcaatatg ggaacaaatc tatctctacc atgaaagtga tgcagtttca gggaatgaag
cacagagcca gtcccccat ctccctgcct ccggtcacac accttgacct cacccccagc cctgacgtgc ccctgacct
cttgaagagg aagctgctga gaaccaacga cgtgaaggaa tcccagaatc tcattgggca gatccagcaa ttcttgatg
ccaggcacgt cattgagaag tctgtgcaca agatcgttc cctgctggcg ggattgggg aaactgctga gag (SEQ ID NO:
10 26)

Murine Legumain anti-sense riboprobe:

Forward Primer:

5'ACACCAACACCAGCCATGTC3' (SEQ ID NO: 27)

Reverse primer (with T7 site):

5'GACTGATAATACGACTCACTATAGGGCGACTCTCAGCAGTTTCCCCAAATC3' (SEQ ID NO: 28)

The probe sequence is 313 nucleotides long and has the following sequence:

ctctcagcag ttccccaaa tcccgccagc agggaaacga tcttgtgcac agacttctca atgacgtgcc tggcatccag
aaattgctgg atctgccc aa tgagattctg ggattccttc acgtcgttgg ttctcagcag ctctctcttc aagatggctca
20 ggggcacgtc agggctgggg gtgaggtaa ggtgtgtgac cggaggcagg gagatggggg aactggctct gtgcttcatt
ccctgaaact gcatcacttt catggtagag atagatttgt tcccatattg catgacatgg ctggtgttgg tgt (SEQ ID NO: 29)

Sections of paraffin embedded tissue were de-paraffinized with xylene, 2 changes, 3 minutes each, and rehydrated to water. After a rinse in RNase-free water and phosphate buffered saline (PBS), permeabilization was performed by incubation with 0.2% Triton-X 100/PBS for 15
25 minutes. After 2 washes with PBS, each at 3 minutes, the sections were ready for proteinase K (PK)(Sigma) treatment. Sections were immersed in 0.1M Tris and 50mM EDTA (Sigma) (pH 8.0) pre-warmed at 37 °C containing 5 µg/ml PK for 15 minutes. PK activities were stopped by 0.1M glycine/PBS for 5 minutes followed by a post fixation with 4% paraformaldehyde for 3 minutes and PBS rinsed. To prevent non-specific electrostatic binding of the probe, sections
30 were immersed in 0.25% acetic anhydride and 0.1M triethanolamine solution (pH 8.0) for 10 minutes, followed by 15 seconds in 20% acetic acid at 4 °C. After 3 changes in PBS, 5 minutes

each, sections were dehydrated through 70%, 90% and 100% ethanol, each at 3 minutes. The sections were completely air dried before 40 µl of pre-hybridization buffer was applied, covered with Parafilm and incubated at 52 °C for 30 minutes to reduce non-specific binding. Parafilm was removed and 40 µl of hybridization buffer containing 5 ng/µl of digoxigenin-labeled probe was applied to each section, recovered with Parafilm and incubated overnight at 52 °C.

Parafilm was carefully removed and sections were immersed into 2x saline sodium citrate (SSC) (Sigma)/0.1% lauryl sulphate (SDS) (Sigma) at room temperature, 4 changes, 5 minutes each. To ensure specific binding of the probe, sections were washed in a high stringency solution containing 0.1x SSC/0.1% SDS at 52 °C, 2 changes, 10 minutes each. Endogenous peroxidase was quenched by immersion of sections in 3% H₂O₂, 15 minutes at room temperature followed by 3 washes in PBS, 2 minutes each. The labeled probe was detected with anti-digoxigenin antibody conjugated to peroxidase complex (Roche) diluted 1:50 in 2% normal sheep serum/0.1% Triton X-100. Labeled probe was developed with DAB (Vector Laboratories), washed in water, stained briefly with hematoxylin, dehydrate in graded alcohol and mounted in Permount mountant (Fisher Scientific) before microscopic examination.

The results indicate that no staining was observed in any of the paws treated with the sense probes (negative control). Expression of all of the genes described below was detected in joints of animals with collagen induced arthritis. No staining was seen in untreated animals.

More particularly, individual cells expressing the RNAs tested for were identified. FST1 mRNA positive cells were neutrophils, macrophages, fibroblasts, osteoblasts. No FST1 was found in bone tissue. SAA-3 mRNA positive cells were, neutrophils, macrophages, fibroblasts, superficial epidermis and chondrocytes. No staining with SAA-3 was seen in the articular cartilage. SLP-1 mRNA positive cells were osteoblasts, fibroblast and a focal area of chondrogenesis. The macrophages seemed to be positive (mild), endothelial cells appeared to be positive (mild) and neutrophils seemed to be negative for SLP-1 mRNA. Lymphocytes were difficult to identify in the SLP-1 hybridized sections. Legumain mRNA positive cells were seen in the epidermis. Osteoblasts and fibroblast had positive cytoplasmic staining with the Legumain antisense probe. The macrophages were positive (mild) and neutrophils appeared to be negative for Legumain mRNA. Lymphocytes were difficult to identify in the Legumain hybridized sections.

BOOK REVIEW

5